

73  
DEX - 0.312

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 August 2002 (29.08.2002)

PCT

(10) International Publication Number  
WO 02/066605 A2

(51) International Patent Classification: C12N

San Francisco, CA 94081 (US). SUN, Yongming [CN/US];  
551 Shoal Drive, Redwood City, CA 94065 (US). LIU,  
Chenghua [CN/US]; 1125 Ranchero Way, #14, San Jose,  
CA 95117 (US).

(21) International Application Number: PCT/US02/04284

(22) International Filing Date: 14 February 2002 (14.02.2002)

(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrrell  
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
(60/268,999) 15 February 2001 (15.02.2001) US

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VN, YU, ZA, ZM, ZW.

(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier application:  
US 60/268,999 (CIP)  
Filed on 15 February 2001 (15.02.2001)

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

(71) Applicant (for all designated States except US): DI-  
ADENUS, INC. [US/US]; 343 Oyster Point Boulevard,  
South San Francisco, CA 94080 (US).

(72) Inventors: and

(75) Inventors/Applicants (for US only): SALCEDA, Susana  
[AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136  
(US). MACINA, Roberto, A. [AR/US]; 4118 Crescendo  
Avenue, San Jose, CA 95136 (US). HU, Ping [US/US]; 108  
Palmer Street, San Ramon, CA 94583 (US). RECIPON,  
Herve [FR/US]; 85 Fortuna Avenue, San Francisco, CA  
94115 (US). KARRA, Kalpana [IN/US]; 6511 Rainbow  
Drive, San Jose, CA 95129 (US). CAFFERKEY, Robert  
[IE/US]; 849 West Orange Avenue, Apartment 1030, South

**Published:**

without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.



WO 02/066605 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue, identifying breast tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.



14

-1-

## COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application  
5 Serial No. 60/268,999 filed February 15, 2001, which is herein incorporated by reference  
in its entirety.

### FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and  
polypeptides present in normal and neoplastic breast cells, including fragments, variants  
10 and derivatives of the nucleic acids and polypeptides. The present invention also relates  
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of  
the polypeptides of the invention. The invention also relates to compositions comprising  
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists  
of the invention and methods for the use of these compositions. These uses include  
15 identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-  
cancerous disease states in breast tissue, identifying breast tissue and monitoring and  
identifying and/or designing agonists and antagonists of polypeptides of the invention.  
The uses also include gene therapy, production of transgenic animals and cells, and  
production of engineered breast tissue for treatment and research.

### 20 BACKGROUND OF THE INVENTION

Excluding skin cancer, breast cancer, also called mammary tumor, is the most  
common cancer among women, accounting for a third of the cancers diagnosed in the  
United States. One in nine women will develop breast cancer in her lifetime and about  
192,000 new cases of breast cancer are diagnosed annually with about 42,000 deaths.  
25 Bevers, *Primary Prevention of Breast Cancer*, in BREAST CANCER, 20-54 (Kelly K Hunt  
et al., ed., 2001); Kochanek et al., 49 Nat'l. Vital Statistics Reports 1, 14 (2001).

In the treatment of breast cancer, there is considerable emphasis on detection and  
risk assessment because early and accurate staging of breast cancer has a significant  
impact on survival. For example, breast cancer detected at an early stage (stage T0,  
30 discussed below) has a five-year survival rate of 92%. Conversely, if the cancer is not  
detected until a late stage (i.e., stage T4), the five-year survival rate is reduced to 13%.  
AJCC Cancer Staging Handbook pp. 164-65 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998).  
Some detection techniques, such as mammography and biopsy, involve increased

-2-

discomfort, expense, and/or radiation, and are only prescribed only to patients with an increased risk of breast cancer.

Current methods for predicting or detecting breast cancer risk are not optimal. One method for predicting the relative risk of breast cancer is by examining a patient's risk factors and pursuing aggressive diagnostic and treatment regimens for high risk patients. A patient's risk of breast cancer has been positively associated with increasing age, nulliparity, family history of breast cancer, personal history of breast cancer, early menarche, late menopause, late age of first full term pregnancy, prior proliferative breast disease, irradiation of the breast at an early age and a personal history of malignancy. Lifestyle factors such as fat consumption, alcohol consumption, education, and socioeconomic status have also been associated with an increased incidence of breast cancer although a direct cause and effect relationship has not been established. While these risk factors are statistically significant, their weak association with breast cancer limited their usefulness. Most women who develop breast cancer have none of the risk factors listed above, other than the risk that comes with growing older. NIH Publication No. 00-1556 (2000).

Current screening methods for detecting cancer, such as breast self exam, ultrasound, and mammography have drawbacks that reduce their effectiveness or prevent their widespread adoption. Breast self exams, while useful, are unreliable for the detection of breast cancer in the initial stages where the tumor is small and difficult to detect by palpitation. Ultrasound measurements require skilled operators at an increased expense. Mammography, while sensitive, is subject to over diagnosis in the detection of lesions that have questionable malignant potential. There is also the fear of the radiation used in mammography because prior chest radiation is a factor associated with an increase incidence of breast cancer.

At this time, there are no adequate methods of breast cancer prevention. The current methods of breast cancer prevention involve prophylactic mastectomy (mastectomy performed before cancer diagnosis) and chemoprevention (chemotherapy before cancer diagnosis) which are drastic measures that limit their adoption even among women with increased risk of breast cancer. Bevers, *supra*.

A number of genetic markers have been associated with breast cancer. Examples of these markers include carcinoembryonic antigen (CEA) (Mughal et al., 249 JAMA 1881 (1983)) MUC-1 (Frische and Liu, 22 J. Clin. Ligand 320 (2000)), HER-2/neu (Haris et al., 15 Proc.Am.Soc.Clin.Oncology. A96 (1996)), uPA, PAI-1, LPA, LPC,



-3-

RAK and BRCA (Esteva and Fritsche, *Serum and Tissue Markers for Breast Cancer*, in BREAST CANCER, 286-308 (2001)). These markers have problems with limited sensitivity, low correlation, and false negatives which limit their use for initial diagnosis. For example, while the BRCA1 gene mutation is useful as an indicator of an increased risk for breast cancer, it has limited use in cancer diagnosis because only 6.2 % of breast cancers are BRCA1 positive. Malone et al., 279 JAMA 922 (1998). See also, Mewman et al., 279 JAMA 915 (1998) (correlation of only 3.3%).

Breast cancers are diagnosed into the appropriate stage categories recognizing that different treatments are more effective for different stages of cancer. Stage TX indicates that primary tumor cannot be assessed (i.e., tumor was removed or breast tissue was removed). Stage T0 is characterized by abnormalities such as hyperplasia but with no evidence of primary tumor. Stage Tis is characterized by carcinoma in situ, intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor. Stage T1 is characterized as having a tumor of 2 cm or less in the greatest dimension. Within stage T1, Tmic indicates microinvasion of 0.1 cm or less, T1a indicates a tumor of between 0.1 to 0.5 cm, T1b indicates a tumor of between 0.5 to 1 cm, and T1c indicates tumors of between 1 cm to 2 cm. Stage T2 is characterized by tumors from 2 cm to 5 cm in the greatest dimension. Tumors greater than 5 cm in size are classified as stage T4. Within stage T4, T4a indicates extension of the tumor to the chest wall, T4b indicates edema or ulceration of the skin of the breast or satellite skin nodules confined to the same breast, T4c indicates a combination of T4a and T4b, and T4d indicates inflammatory carcinoma. AJCC Cancer Staging Handbook pp. 159-70 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998). In addition to standard staging, breast tumors may be classified according to their estrogen receptor and progesterone receptor protein status. Fisher et al., 7 Breast Cancer Research and Treatment 147 (1986). Additional pathological status, such as HER2/neu status may also be useful. Thor et al., 90 J.Nat'l.Cancer Inst. 1346 (1998); Paik et al., 90 J.Nat'l.Cancer Inst. 1361 (1998); Hutchins et al., 17 Proc.Am.Soc.Clin.Oncology A2 (1998).; and Simpson et al., 18 J.Clin.Oncology 2059 (2000).

In addition to the staging of the primary tumor, breast cancer metastases to regional lymph nodes may be staged. Stage NX indicates that the lymph nodes cannot be assessed (e.g., previously removed). Stage N0 indicates no regional lymph node metastasis. Stage N1 indicates metastasis to movable ipsilateral axillary lymph nodes. Stage N2 indicates metastasis to ipsilateral axillary lymph nodes fixed to one another or

-4-

to other structures. Stage N3 indicates metastasis to ipsilateral internal mammary lymph nodes. Id.

Stage determination has potential prognostic value and provides criteria for designing optimal therapy. Simpson et al., 18 J. Clin. Oncology 2059 (2000). Generally, 5 pathological staging of breast cancer is preferable to clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred if it were as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of breast cancer would be improved by detecting new markers in cells, tissues, or bodily fluids which could differentiate 10 between different stages of invasion. Progress in this field will allow more rapid and reliable method for treating breast cancer patients.

Treatment of breast cancer is generally decided after an accurate staging of the primary tumor. Primary treatment options include breast conserving therapy (lumpectomy, breast irradiation, and surgical staging of the axilla), and modified radical 15 mastectomy. Additional treatments include chemotherapy, regional irradiation, and, in extreme cases, terminating estrogen production by ovarian ablation.

Until recently, the customary treatment for all breast cancer was mastectomy. Fonseca et al., 127 Annals of Internal Medicine 1013 (1997). However, recent data indicate that less radical procedures may be equally effective, in terms of survival, for 20 early stage breast cancer. Fisher et al., 16 J. of Clinical Oncology 441 (1998). The treatment options for a patient with early stage breast cancer (i.e., stage Tis) may be breast-sparing surgery followed by localized radiation therapy at the breast. Alternatively, mastectomy optionally coupled with radiation or breast reconstruction may be employed. These treatment methods are equally effective in the early stages of breast 25 cancer.

Patients with stage I and stage II breast cancer require surgery with chemotherapy and/or hormonal therapy. Surgery is of limited use in Stage III and stage IV patients. Thus, these patients are better candidates for chemotherapy and radiation therapy with surgery limited to biopsy to permit initial staging or subsequent restaging because cancer 30 is rarely curative at this stage of the disease. AJCC Cancer Staging Handbook 84, ¶. 164-65 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998).

In an effort to provide more treatment options to patients, efforts are underway to define an earlier stage of breast cancer with low recurrence which could be treated with lumpectomy without postoperative radiation treatment. While a number of attempts have

-5-

been made to classify early stage breast cancer, no consensus recommendation on postoperative radiation treatment has been obtained from these studies. Page et al., 75 Cancer 1219 (1995); Fisher et al., 75 Cancer 1223 (1995); Silverstein et al., 77 Cancer 2267 (1996).

5 As discussed above, each of the methods for diagnosing and staging breast cancer is limited by the technology employed. Accordingly, there is need for sensitive molecular and cellular markers for the detection of breast cancer. There is a need for molecular markers for the accurate staging, including clinical and pathological staging, of breast cancers to optimize treatment methods. Finally, there is a need for sensitive  
10 molecular and cellular markers to monitor the progress of cancer treatments, including markers that can detect recurrence of breast cancers following remission.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while  
15 indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

## 20 SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast; identify and monitor breast tissue; and identify  
25 and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered breast tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to breast cells and/or breast tissue. These breast specific nucleic acids  
30 (BSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the BSNA is genomic DNA, then the BSNA is a breast specific gene (BSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast. In a more preferred embodiment, the nucleic acid molecule encodes a

-6-

polypeptide that comprises an amino acid sequence of SEQ ID NO: 116 through 210. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 115. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence

5 similarity to a nucleic acid molecule encoding a BSP, or that selectively hybridize or exhibit substantial sequence similarity to a BSNA, as well as allelic variants of a nucleic acid molecule encoding a BSP, and allelic variants of a BSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a BSP or that comprises a part of a nucleic acid sequence of a BSNA are also provided.

10 A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a BSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment  
15 of a BSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a BSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a BSNA.

20 Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a BSP.

25 The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a BSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

30 Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a

-7-

preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring breast tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered breast tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast. The invention provides methods of using the polypeptides of the invention to identify and/or monitor breast tissue, and to produce engineered breast tissue.

The agonists and antagonists of the instant invention may be used to treat breast cancer and non-cancerous disease states in breast and to produce engineered breast tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

## DETAILED DESCRIPTION OF THE INVENTION

### Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, e.g., Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor

-8-

- Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4<sup>th</sup> Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

- The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages

-9-

(*e.g.*, alpha anomeric nucleic acids, etc.) The term “nucleic acid molecule” also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated  
5 sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A “gene” is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround  
10 the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term “exon” refers to a  
15 nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term “intron” refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be “spliced out” during processing of the transcript.

20 A nucleic acid molecule or polypeptide is “derived” from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An “isolated” or “substantially pure” nucleic acid or polynucleotide (*e.g.*, an  
25 RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a  
30 portion of a polynucleotide in which the “isolated polynucleotide” is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term “isolated” or “substantially pure” also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized

-10-

polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant  
5 vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to  
10 occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression  
15 or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule,  
20 or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally  
25 comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35,  
30 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.



-11-

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See *e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the

-12-

polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

- 5 In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

- The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at
- 10 least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using
- 15 FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000);
- 20 Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or
- 25 using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

- A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its
- 30 complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"

-13-

interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point ( $T_m$ ) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the  $T_m$  for the specific DNA hybrid under a particular set of conditions. The  $T_m$  is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

-14-

The  $T_m$  for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where  $l$  is the length of the hybrid in base pairs.

The  $T_m$  for a particular RNA-RNA hybrid can be estimated by the formula:

5 
$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The  $T_m$  for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

10 In general, the  $T_m$  decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C  
15 would be subtracted from the calculated  $T_m$  of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

20 An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without  
25 formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a  
30 library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide

-15-

concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See Sambrook et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

- 5 Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An
- 10 exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

- As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode
- 15 polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

- Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:
- 20  $T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/\text{N})$ ,  
wherein N is change length and the  $[\text{Na}^+]$  is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the  $T_m$ ) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using
- 25 mismatched probes, pools of degenerate probes or “guessmers,” as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

- The term “digestion” or “digestion of DNA” refers to catalytic cleavage of the
- 30 DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the

-16-

purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical

-17-

Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are  
5 disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration  
10 may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a BSP or is a BSNA. The nucleic acid molecule may be mutated by any method known in the art  
15 including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-  
20 33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a  
25 PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA  
30 molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

-18-

The term “*in vivo* mutagenesis” refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These “mutator” strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term “cassette mutagenesis” refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence:

The term “recursive ensemble mutagenesis” refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term “exponential ensemble mutagenesis” refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that



-19-

enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and

5 transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule

10 capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral

15 genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable

20 of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector.

25 However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell

30 but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

-20-

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a BSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,

-21-

followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides

-22-

are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences.

- 5 Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be
- 10 produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

- 15 The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the
- 20 same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

- 25 The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce
- 30 an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH<sub>2</sub>NH--, --CH<sub>2</sub>S--, --CH<sub>2</sub>-CH<sub>2</sub>--,  
--CH=CH--(cis and trans), --COCH<sub>2</sub>--, --CH(OH)CH<sub>2</sub>--, and --CH<sub>2</sub>SO--, by methods

-23-

well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (e.g., D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not

-24-

substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are

5 described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations

10 follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2<sup>nd</sup> Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as  $\alpha$ -,  $\alpha$ -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention.

15 Examples of unconventional amino acids include: 4-hydroxyproline,  $\gamma$ -carboxyglutamate,  $\epsilon$ -N,N,N-trimethyllysine,  $\epsilon$ -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand

20 direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological

25 activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have

30 similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90%

-25-

sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. See, e.g., Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from

-26-

different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.,* GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer  
5 program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

|    |                       |               |
|----|-----------------------|---------------|
|    | Expectation value:    | 10 (default)  |
|    | Filter:               | seg (default) |
| 10 | Cost to open a gap:   | 11 (default)  |
|    | Cost to extend a gap: | 1 (default)   |
|    | Max. alignments:      | 100 (default) |
|    | Word size:            | 11 (default)  |
|    | No. of descriptions:  | 100 (default) |
| 15 | Penalty Matrix:       | BLOSUM62      |

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number  
20 of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.,* FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best  
25 overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

30 An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.,* a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')<sub>2</sub>, Fv,



-27-

dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')<sub>2</sub> fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature 341: 544-546 (1989).*

10 By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

15 A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science 242: 423-426 (1988); Huston et al., Proc. Natl. Acad. Sci. USA 85: 5879-5883 (1988).* Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA 90: 6444-6448 (1993); Poljak et al., Structure 2: 1121-1123 (1994).* One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

30 An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-

-28-

chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

10 A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1  $\mu$ M, preferably less than 100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "breast specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the breast as compared to other tissues in the body. In a preferred embodiment, a "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide

levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

5

*Nucleic Acid Molecules*

One aspect of the invention provides isolated nucleic acid molecules that are specific to the breast or to breast cells or tissue or that are derived from such nucleic acid molecules. These isolated breast specific nucleic acids (BSNAs) may comprise a cDNA,  
10 a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast, a breast-specific polypeptide (BSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 116 through 210. In  
15 another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 115.

A BSNA may be derived from a human or from another animal. In a preferred embodiment, the BSNA is derived from a human or other mammal. In a more preferred embodiment, the BSNA is derived from a human or other primate. In an even more  
20 preferred embodiment, the BSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a BSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a BSP. However, in a  
25 preferred embodiment, the hybridizing nucleic acid molecule encodes a BSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 116 through 210. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic  
30 acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 115.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under moderate stringency conditions. In a more preferred

-30-

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 116 through 210. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 115. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a BSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human BSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 210. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a BSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 210, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a BSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a BSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a BSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a BSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 115, more preferably at least 70%, even more preferably at least 80% and even more

-31-

preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a BSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a BSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a BSNA or to a nucleic acid molecule encoding a BSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the BSNA or the nucleic acid molecule encoding a BSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 116 through 210 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 115. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the BSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a BSNA. Further, the substantially similar nucleic acid molecule may or may not be a BSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a BSNA.

-32-

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a BSNA or a nucleic acid encoding a BSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a BSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a BSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 115. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a BSP. However, in a preferred embodiment, the part encodes a BSP. In one aspect, the invention comprises a part of a BSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a BSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a BSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a BSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

-33-

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with  $^{33}\text{P}$ ,  $^{32}\text{P}$ , and  $^{35}\text{S}$ , such as  $\alpha$ - $^{32}\text{P}$ -dATP,  $\alpha$ - $^{32}\text{P}$ -dCTP,  $\alpha$ - $^{32}\text{P}$ -dGTP,  $\alpha$ - $^{32}\text{P}$ -dTTP,  $\alpha$ - $^{32}\text{P}$ -3'dATP,  $\alpha$ - $^{32}\text{P}$ -ATP,  $\alpha$ - $^{32}\text{P}$ -CTP,  $\alpha$ - $^{32}\text{P}$ -GTP,  $\alpha$ - $^{32}\text{P}$ -UTP,  $\alpha$ - $^{35}\text{S}$ -dATP,  $\alpha$ - $^{35}\text{S}$ -GTP,  $\alpha$ - $^{33}\text{P}$ -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade

-34-

Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994);



-35-

Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a  
5 free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report  
10 specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.*, Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517;  
15 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more  
20 native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd  
25 (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation,  
30 phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having

-36-

normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'.

Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808;

5 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019;  
5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233;  
5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253;  
5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by  
10 reference in their entireties. In a preferred embodiment, the modified internucleoside  
linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include

15 those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones;  
20 and others having mixed N, O, S and CH<sub>2</sub> component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562;  
5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307;  
5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289;  
25 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced  
30 with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S.

-37-

patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 5 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The  $T_m$  of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the  $T_m$  of the 10 corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed 15 a PNA/DNA 15-mer lowers the  $T_m$  by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the  $T_m$  by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases 20 do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g., Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

25 Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, 30 Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly

-38-

comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001);  
5 Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201  
10 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

*Methods for Using Nucleic Acid Molecules as Probes and Primers*

The isolated nucleic acid molecules of the present invention can be used as  
15 hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

20 In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a BSNA, such as deletions, insertions, translocations, and duplications of the BSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and  
25 Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that  
30 include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify BSNA in, and

-39-

isolate BSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A<sup>+</sup>-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.,* Schwarczacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to BSNA, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a BSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 210. In another preferred embodiment, the probe or primer is derived from a BSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer

-40-

in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.,* Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.,* in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.,* U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.,* Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by

-41-

reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. See, e.g., Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate  
5 either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to  
10 a porous substrate, e.g., a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, e.g., a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without  
15 limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including  
20 rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or  
25 some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, e.g. on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such  
30 as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

-42-

*Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides*

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

5       The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids  
10 of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones  
15 *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000);  
20 Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an  
25 expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part  
30 of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for



-43-

example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred  
5 embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their  
10 derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989,  $\lambda$ GT10 and  $\lambda$ GT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline,  
15 chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous  
20 recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and  
25 a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 $\mu$  plasmids and derivatives thereof, and improved shuttle vectors such as those  
30 described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

-44-

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors  
5 are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the  
10 polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway.  
15 Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, e.g., in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A).  
20 Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in  
25 mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (e.g., vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (e.g., bovine papillomavirus),  
30 and retroviral vectors (e.g., murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

-45-

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast  $\alpha$ -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

-46-

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the BSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit  $\beta$ -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (e.g., ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two

-47-

copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

-48-

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or  
5 identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III  
10 protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996).  
15 Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the  $\alpha$ -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet  
20 derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring  
25 proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic  
30 fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of

-49-

protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (see, e.g. Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (see, e.g. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (see, e.g. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (see, e.g. Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. See also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and

-50-

cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent  
5 antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-  
10 293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally  
15 well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must  
20 be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into  
25 the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide BSPs with such post-  
30 translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and



-51-

racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational  
5 modifications. See, e.g., [www.expasy.org](http://www.expasy.org) (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-  
10 anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in  
15 web sites such as the Delta Mass database <http://www.abrf.org/ABRF/ResearchCommittees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and  
<http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a  
20 revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html>  
25 (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from  
30 normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another

common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue.

- 5 Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

- Another post-translational modification that may be altered in cancer cells is
- 10 prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).
- 15

- Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to
- 20 the corresponding polypeptides from noncancerous cells.

- Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is
- 25 cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell
- 30 compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method

-53-

known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website [www.expasy.org](http://www.expasy.org). The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational

-54-

modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its  
5 controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture  
10 requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention  
15 may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in*  
20 *vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell  
25 (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See, for instance, Ausubel, supra, and Sambrook et al., supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an  
30 expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be

-55-

able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from breast are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human breast cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

- 5 Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with  $\text{CaCl}_2$ , or a solution of  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Rb}^+$  or  $\text{K}^+$ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent
- 10 strains are also available commercially (e.g., Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse
- 15 treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) ([http://www.biorad.com/LifeScience/pdf/](http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf) New\_Gene\_Pulser.pdf).

- Vectors can be introduced into yeast cells by spheroplasting, treatment with
- 20 lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol
- 25 (PEG) and  $\text{Ca}^{2+}$ . Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

- For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of
- 30 PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

-57-

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991).

- 5 The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

- Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be
- 10 coprecipitated with  $\text{CaPO}_4$  or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for  $\text{CaPO}_4$  transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN®
- 15 Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) ([http://www.bio-rad.com/LifeScience/pdf/](http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)
- 20 [New\\_Gene\\_Pulser.pdf](http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. See, e.g., Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA*
- 25 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- Purification of recombinantly expressed proteins is now well by those skilled in the art. See, e.g., Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid
- 30 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.),

-58-

Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an  
5 expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

10 Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a breast specific polypeptide (BSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ  
15 ID NO: 116 through 210. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide,  
20 wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a BSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 116 through 210. A polypeptide that comprises only a fragment of an entire BSP may or may not be a polypeptide that is also a BSP. For instance, a full-length polypeptide may be  
25 breast-specific, while a fragment thereof may be found in other tissues as well as in breast. A polypeptide that is not a BSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-BSP antibodies. However, in a preferred embodiment, the part or fragment is a BSP. Methods of determining whether a polypeptide is a BSP are  
30 described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of



-59-

which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

5           Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. See, e.g., Lerner, *Nature* 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further  
10 described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

          Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as  
15 competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their  
20 entireties.

          The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids,  
25 or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

          One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, e.g., a BSNA, encoding the polypeptide and then  
30 expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. See, e.g., Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992),

-60-

*supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably a BSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a  
5 fragment of the polypeptide, preferably a BSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared  
10 to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be breast-specific. In a preferred embodiment, the mutein is breast-specific. In a preferred  
15 embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 116 through 210. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to  
20 a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210.

25 A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical  
30 techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino

-61-

acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is breast-specific, as described below. Multiple random mutations can be introduced into the

5 gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*,

10 Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the

15 polypeptide is homologous to a BSP. In an even more preferred embodiment, the polypeptide is homologous to a BSP selected from the group having an amino acid sequence of SEQ ID NO: 116 through 210. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a BSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity

20 to an comprising an amino acid sequence of SEQ ID NO: 116 through 210. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210. In a yet more

25 preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%,

30 99.7%, 99.8% or 99.9% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a BSNA. In a preferred embodiment,

-62-

the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a BSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSNA is selected from the group consisting of SEQ ID NO: 1 through 115. In another preferred embodiment, the

- 5 homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a BSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSP is selected from the group consisting of SEQ ID NO: 116 through 210.

- The homologous polypeptide may be a naturally-occurring one that is derived
- 10 from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 116 through 210. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the BSP is a member of a family of
- 15 polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The
- 20 naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and
- 25 subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a BSP. Further, the homologous protein may or may not encode polypeptide that is a BSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a BSP.
- 30 Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the

-63-

binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a BSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 210. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 115.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 210, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

-64-

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), e.g., offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

-65-

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SAMP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SAMP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-BSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred

-66-

embodiment, the polypeptide is a BSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 116 through 210. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds  
5 compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a BSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH<sub>2</sub>NH--, --CH<sub>2</sub>S--, --CH<sub>2</sub>-CH<sub>2</sub>--,  
--CH=CH--(cis and trans), --COCH<sub>2</sub>--, --CH(OH)CH<sub>2</sub>-- and --CH<sub>2</sub>SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids  
10 of a BSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly  
15 added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Krole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical  
20 synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford  
25 Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added  
30 using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and tBOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during



-67-

synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyI quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding tBOC derivative (both from  
5 Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl  
10 side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially,  
15 including, e.g., Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-  
20 trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-  
25 aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- $\beta$ -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-  
30 hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtioic acid, Fmoc-D,L-3-amino-3-

-68-

phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

#### *Fusion Proteins*

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 116 through 210, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 115, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

-69-

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.,* Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His<sup>6</sup> tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. *See also* the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See Bartel et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et*

- al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbri *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entirety. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins,  $\beta$ -galactosidase, biotin *trpE*, protein A,  $\beta$ -lactamase,  $\alpha$ -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), *lacZ*, green fluorescent protein (GFP), yeast  $\alpha$  mating factor, GAL4 transcription activation or DNA binding domain,

-71-

luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art.

- 5 Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding  
10 proteins or other molecules that bind to the BSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize BSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to  
15 assay for the polypeptides of the present invention, particularly BSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of BSPs, as for example by immunoprecipitation, and for use  
20 as specific agonists or antagonists of BSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine  
25 scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are  
30 available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of

-72-

one having ordinary skill in the art. See, e.g., Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, e.g., by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated  
5 proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the  
10 present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the  
15 present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the  
20 present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, e.g. in serum, that bind  
25 specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, e.g. in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include  
30 polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

-73-

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with  
5 sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the  
10 surface-bound protein to indicate biological interaction there between.

#### Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid  
15 molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a BSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 116 through 210, or a fragment, mutein,  
20 derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also  
25 due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a BSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a BSP may be indicative of cancer. Differential degradation of the C or N-terminus of a BSP may also be a marker or target for anticancer therapy. For example, a BSP may  
30 be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention

-74-

will discriminate over adventitious binding to non-BSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human breast.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about  $1 \times 10^{-6}$  molar (M), typically at least about  $5 \times 10^{-7}$  M,  $1 \times 10^{-7}$  M, with affinities and avidities of at least  $1 \times 10^{-8}$  M,  $5 \times 10^{-9}$  M,  $1 \times 10^{-10}$  M and up to  $1 \times 10^{-13}$  M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered



-75-

antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but  
5 also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are hereby incorporated in their entirety. In such cases, as with the transgenic human-  
10 antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when  
15 conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention  
20 can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

25 Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From  
30 Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

-76-

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entirety, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: e.g., genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein

-77-

(pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.,* Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997);

5 Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).

Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas

10 (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length

15 antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al.*, *Biosci.*

20 *Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein

25 by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997);

30 and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000);

-78-

Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein  
5 by reference in their entirety.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures  
10 of which are incorporated herein by reference in their entirety.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated  
15 by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic  
20 animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entirety.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present  
25 invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'<sub>2</sub>, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin.*  
30 *Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively

-79-

inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.,* United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco *et al.*, *Proc. Natl.*

*Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to  
5 provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments  
10 encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

15 Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase,  $\beta$ -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl  
20 phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue  
25 tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide ( $H_2O_2$ ), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate  
30 reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe et al., *Methods Enzymol.* 133: 331-53 (1986); Kricka et al., *J. Immunoassay* 17(1): 67-83

-81-

(1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

5 As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

10 For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

15 Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568,  
20 BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5,  
25 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting  
30 applications, they can usefully be labeled with radioisotopes, such as <sup>33</sup>P, <sup>32</sup>P, <sup>35</sup>S, <sup>3</sup>H, and <sup>125</sup>I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be <sup>228</sup>Th, <sup>227</sup>Ac, <sup>225</sup>Ac, <sup>223</sup>Ra, <sup>213</sup>Bi, <sup>212</sup>Pb,

-82-

$^{212}\text{Bi}$ ,  $^{211}\text{At}$ ,  $^{203}\text{Pb}$ ,  $^{194}\text{Os}$ ,  $^{188}\text{Re}$ ,  $^{186}\text{Re}$ ,  $^{153}\text{Sm}$ ,  $^{149}\text{Tb}$ ,  $^{131}\text{I}$ ,  $^{125}\text{I}$ ,  $^{111}\text{In}$ ,  $^{105}\text{Rh}$ ,  $^{99\text{m}}\text{Tc}$ ,  $^{97}\text{Ru}$ ,  $^{90}\text{Y}$ ,  $^{90}\text{Sr}$ ,  $^{88}\text{Y}$ ,  $^{72}\text{Se}$ ,  $^{67}\text{Cu}$ , or  $^{47}\text{Sc}$ .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI

5 contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives  
10 thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000);  
15 and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention,  
20 to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

25 Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to  
30 paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.



-83-

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the

5 antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by

10 one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the

15 antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

#### Transgenic Animals and Cells

20

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a BSP. In a preferred embodiment, the BSP comprises an amino acid

25 sequence selected from SEQ ID NO: 116 through 210, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a BSNA of the invention, preferably a BSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 115, or a part, substantially similar nucleic acid molecule, allelic variant

30 or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human BSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric

- homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson et al., Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).*

- Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g., Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology 11: 1263-1270 (1993); Wright et al., Biotechnology 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989) retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol. 3: 1803-1814 (1983); introduction using a gene gun (*see, e.g., Ulmer et al., Science 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g., Lavitrano et al., Cell 57: 717-723 (1989)).******

- Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g., Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810-813 (1997)).* The present invention provides for transgenic animals that carry the transgene (*i.e., a nucleic acid molecule of the invention*) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i. e., mosaic animals or chimeric animals.*

- The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, *e. g., head-to-head tandems or head-to-tail tandems.* The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g., the teaching of Lasko et al. et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992).* The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to

-85-

verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. See, e.g., Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. See, e.g., Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable  
5 marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications  
10 to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and *Thompson, supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

15 In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells,  
20 blood cells (*e.g., lymphocytes*), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g., by transduction* (using viral vectors, and preferably  
25 vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve  
30 expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g., in the circulation*, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g., genetically engineered fibroblasts* can be implanted as part of a skin graft;

-87-

genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

#### 15 Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 115 and SEQ ID NO: 116 through 210 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable

-89-

medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

#### Diagnostic Methods for Breast Cancer

10

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a BSNA or a BSP in a human patient that has or may have breast cancer, or who is at risk of developing breast cancer, with the expression of a BSNA or a BSP in a normal human control. For purposes of the present invention, "expression of a BSNA" or "BSNA expression" means the quantity of BSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a BSP" or "BSP expression" means the amount of BSP that can be measured by any method known in the art or the level of translation of a BSG BSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing breast cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of BSNA or BSP in cells, tissues, organs or bodily fluids compared with levels of BSNA or BSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a BSNA or BSP in the patient versus the normal human control is associated with the presence of breast cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in the structure of the mRNA of a BSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in a BSP compared to a BSP from a normal control. These

-90-

changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the BSP or subcellular BSP localization.

In a preferred embodiment, the expression of a BSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 116 through 210, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the BSNA expression that is measured is the level of expression of a BSNA mRNA selected from SEQ ID NO: 1 through 115, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. BSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. BSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a BSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, BSNA expression may be compared to a known control, such as normal breast nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a BSP is measured by determining the level of a BSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 210, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of BSNA or BSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of breast cancer. The expression level of a BSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the BSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.*, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the BSP



-91-

structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

5           In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a BSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-BSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the  
10 antibody on the solid support under conditions in which the BSP will bind to the anti-BSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-BSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the BSP to the labeled antibody will occur. After  
15 binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a BSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

20           Other methods to measure BSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-BSP antibody is attached to a solid support and an allocated amount of a labeled BSP and a sample of interest are incubated with the solid support. The amount of labeled BSP detected which is attached to the solid support can be correlated to the quantity of a BSP in the sample.

25           Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is  
30 perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a BSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example,

5 reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

10 Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more BSNAs of interest. In this approach, all or a portion of one or more BSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the  
15 RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid  
20 molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any  
25 other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of BSNA or BSP includes, without limitation, breast tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, breast cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment,  
30 especially when metastasis of a primary breast cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy,

-93-

bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in BSNA or BSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

- 5 All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a BSNA or BSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other BSNA or BSPs as disclosed herein.
- 10 Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular BSNA or BSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more
- 15 preferably at least ten additional cancer markers are used.

#### *Diagnosing*

- In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more BSNA and/or BSPs in a sample from a patient suspected of having breast cancer. In general, the method comprises the steps
- 20 of obtaining the sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP and then ascertaining whether the patient has breast cancer from the expression level of the BSNA or BSP. In general, if high expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times
- 25 higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least
- 30 five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

-94-

The present invention also provides a method of determining whether breast cancer has metastasized in a patient. One may identify whether the breast cancer has metastasized by measuring the expression levels and/or structural alterations of one or more BSNA and/or BSPs in a variety of tissues. The presence of a BSNA or BSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (e.g., the same tissue from another individual) is indicative of metastasis if high level expression of a BSNA or BSP is associated with breast cancer. Similarly, the presence of a BSNA or BSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a BSNA or BSP is associated with breast cancer. Further, the presence of a structurally altered BSNA or BSP that is associated with breast cancer is also indicative of metastasis.

In general, if high expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The BSNA or BSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with breast cancers or other breast related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of breast disorders.

#### *Staging*

The invention also provides a method of staging breast cancer in a human patient. The method comprises identifying a human patient having breast cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more BSNA or BSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more BSNA or BSPs is determined for each stage to obtain a

-95-

standard expression level for each BSNA and BSP. Then, the BSNA or BSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The BSNA or BSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the BSNA and BSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a BSNA or BSP to determine the stage of a breast cancer.

#### *Monitoring*

Further provided is a method of monitoring breast cancer in a human patient.

One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the breast cancer. The method comprises identifying a human patient that one wants to monitor for breast cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more BSNA or BSPs, and comparing the BSNA or BSP levels over time to those BSNA or BSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a BSNA or BSP that are associated with breast cancer.

If increased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively.

One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of BSNA or BSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

-96-

Monitoring a patient for onset of breast cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a BSNA and/or BSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more BSNAs and/or BSPs are detected. The presence of higher (or lower) BSNA or BSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly breast cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more BSNAs and/or BSPs of the invention can also be monitored by analyzing levels of expression of the BSNAs and/or BSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

#### *Detection of Genetic Lesions or Mutations*

The methods of the present invention can also be used to detect genetic lesions or mutations in a BSG, thereby determining if a human with the genetic lesion is susceptible to developing breast cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing breast cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the BSGs of this invention, a chromosomal rearrangement of BSG, an aberrant modification of BSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a BSG. Methods to detect such lesions in the BSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

#### Methods of Detecting Noncancerous Breast Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having or known to have a noncancerous breast disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP, comparing the

-97-

expression level or structural alteration of the BSNA or BSP to a normal breast control, and then ascertaining whether the patient has a noncancerous breast disease. In general, if high expression relative to a control of a BSNA or BSP is indicative of a particular noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of a noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a BSNA and/or BSP is associated with a particular noncancerous breast disease by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining which BSNAs and/or BSPs are expressed in the tissue at either a higher or a lower level than in normal breast tissue. In another embodiment, one may determine whether a BSNA or BSP exhibits structural alterations in a particular noncancerous breast disease state by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining the structural alterations in one or more BSNAs and/or BSPs relative to normal breast tissue.

#### Methods for Identifying Breast Tissue

In another aspect, the invention provides methods for identifying breast tissue. These methods are particularly useful in, *e.g.*, forensic science, breast cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is breast tissue or has breast tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising breast tissue or having breast tissue-like characteristics, determining whether the sample expresses one or more BSNAs and/or BSPs, and, if the sample expresses one or more BSNAs and/or BSPs, concluding that the sample comprises breast tissue. In a preferred embodiment, the BSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 116 through 210,

-98-

or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 115, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a BSNA can be accomplished by any method known in the art.

- 5 Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a BSP is expressed. Determining whether a sample expresses a BSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the BSP
- 10 has an amino acid sequence selected from SEQ ID NO: 116 through 210, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two BSNA and/or BSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five BSNA and/or BSPs are determined.
- 15 In one embodiment, the method can be used to determine whether an unknown tissue is breast tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into breast tissue.
- 20 This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new breast tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH,
- 25 aqueous:air interface and various other culture conditions.

#### Methods for Producing and Modifying Breast Tissue

- In another aspect, the invention provides methods for producing engineered breast tissue or cells. In one embodiment, the method comprises the steps of providing cells,
- 30 introducing a BSNA or a BSP into the cells, and growing the cells under conditions in which they exhibit one or more properties of breast tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal breast tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered breast tissue or cells comprises one of these cell types. In another



-99-

embodiment, the engineered breast tissue or cells comprises more than one breast cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the breast cell tissue. Methods for manipulating culture conditions are well-known in the art.

5 Nucleic acid molecules encoding one or more BSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode BSPs having amino acid sequences selected from SEQ ID NO: 116 through 210, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected  
10 from SEQ ID NO: 1 through 115, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a BSP is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial breast tissue may be used to treat patients who have lost some or all of  
15 their breast function.

#### Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives,  
20 antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a BSNA or part thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 115, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity  
25 thereto. In another preferred embodiment, the pharmaceutical composition comprises a BSP or fragment thereof. In a more preferred embodiment, the BSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 116 through 210, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the  
30 pharmaceutical composition comprises an anti-BSP antibody, preferably an antibody that specifically binds to a BSP having an amino acid that is selected from the group consisting of SEQ ID NO: 116 through 210, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

-100-

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in  
5 Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20<sup>th</sup> ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7<sup>th</sup> ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3<sup>rd</sup> ed. (2000), the disclosures of which are incorporated herein by reference in their entireties,  
10 and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation,  
15 topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain  
20 suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium  
25 carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and  
30 alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

-101-

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

5 Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

10 Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or  
15 suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral)  
20 administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and  
25 flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present  
30 invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

-102-

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and  
5 administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate,  
10 isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the  
15 suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as  
20 polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

25 The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops,  
30 tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid

-103-

ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases  
5 as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various  
10 powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for  
15 reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts  
20 tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the  
25 intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example BSP polypeptide, fusion protein, or fragments thereof, antibodies specific for BSP, agonists, antagonists or inhibitors of BSP, which ameliorates the signs or symptoms  
30 of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model

-104-

can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (e.g., 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to

-105-

the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

#### Therapeutic Methods

5           The present invention further provides methods of treating subjects having defects in a gene of the invention, e.g., in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of breast function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any  
10 improvement of a disease, including minor improvements. These methods are discussed below.

#### *Gene Therapy and Vaccines*

          The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be  
15 driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further  
20 described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. See, e.g., Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

25           In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a BSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a BSP are  
30 administered, for example, to complement a deficiency in the native BSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. See, e.g., Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid

-106-

molecule encodes a BSP having the amino acid sequence of SEQ ID NO: 116 through 210, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a BSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in BSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a BSP having the amino acid sequence of SEQ ID NO: 116 through 210, or a fragment, fusion protein, allelic variant or homolog thereof.

#### 10 *Antisense Administration*

Antisense nucleic acid compositions, or vectors that drive expression of a BSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a BSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

15 Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a BSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to BSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

25 Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the BSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.



-107-

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a BSP, preferably a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule  
5 having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

#### *Polypeptide Administration*

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a BSP, a  
10 fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant BSP defect.

Protein compositions are administered, for example, to complement a deficiency in native BSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to BSP. The immune response can  
15 be used to modulate activity of BSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate BSP.

In a preferred embodiment, the polypeptide is a BSP comprising an amino acid  
20 sequence of SEQ ID NO: 116 through 210, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

#### *Antibody, Agonist and Antagonist Administration*

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example,  
30 to antagonize activity of BSP, or to target therapeutic agents to sites of BSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred

-108-

embodiment, the antibody specifically binds to a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

- The present invention also provides methods for identifying modulators which
- 5 bind to a BSP or have a modulatory effect on the expression or activity of a BSP. Modulators which decrease the expression or activity of BSP (antagonists) are believed to be useful in treating breast cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a BSP can
- 10 also be designed, synthesized and tested for use in the imaging and treatment of breast cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the BSPs identified herein. Molecules identified in the library as being capable of binding to a BSP are key candidates for further evaluation for use in the treatment of breast cancer. In a preferred embodiment,
- 15 these molecules will downregulate expression and/or activity of a BSP in cells.

- In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of BSP is administered. Antagonists of BSP can be produced using methods generally known in the art. In particular, purified BSP can be used to screen libraries of pharmaceutical
- 20 agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a BSP.

In other embodiments a pharmaceutical composition comprising an agonist of a BSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

- 25 In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 210, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP encoded by a
- 30 nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

#### *Targeting Breast Tissue*

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the

-109-

breast or to specific cells in the breast. In a preferred embodiment, an anti-BSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if breast tissue needs to be selectively destroyed. This would be useful for targeting and killing breast-cancer cells.

- 5 In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting breast cell function.

In another embodiment, an anti-BSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring breast function, identifying breast cancer tumors,

- 10 and identifying noncancerous breast diseases.

### EXAMPLES

#### Example 1: Gene Expression analysis

BSGs were identified by mRNA subtraction analysis using standard methods. The sequences were extended using GeneBank sequences, Incyte's proprietary database.

- 15 From the nucleotide sequences, predicted amino acid sequences were prepared.

DEX0305\_1, DEX0305\_2 correspond to SEQ ID NO.1, 2 etc. DEX0155 was the parent sequence found in the mRNA subtractions.

|    |            |                 |             |
|----|------------|-----------------|-------------|
|    | DEX0305_1  | DEX0155_1       | DEX0305_116 |
|    | DEX0305_2  | flex DEX0155_1  | DEX0305_117 |
| 20 | DEX0305_3  | DEX0155_2       | DEX0305_118 |
|    | DEX0305_4  | flex DEX0155_2  |             |
|    | DEX0305_5  | DEX0155_3       |             |
|    | DEX0305_6  | DEX0155_4       | DEX0305_119 |
|    | DEX0305_7  | DEX0155_5       |             |
| 25 | DEX0305_8  | DEX0155_6       | DEX0305_120 |
|    | DEX0305_9  | DEX0155_7       | DEX0305_121 |
|    | DEX0305_10 | flex DEX0155_7  |             |
|    | DEX0305_11 | DEX0155_8       | DEX0305_122 |
|    | DEX0305_12 | DEX0155_9       | DEX0305_123 |
| 30 | DEX0305_13 | DEX0155_10      |             |
|    | DEX0305_14 | DEX0155_11      | DEX0305_124 |
|    | DEX0305_15 | DEX0155_12      | DEX0305_125 |
|    | DEX0305_16 | DEX0155_13      |             |
|    | DEX0305_17 | DEX0155_14      | DEX0305_126 |
| 35 | DEX0305_18 | DEX0155_15      | DEX0305_127 |
|    | DEX0305_19 | DEX0155_16      | DEX0305_128 |
|    | DEX0305_20 | DEX0155_17      | DEX0305_129 |
|    | DEX0305_21 | flex DEX0155_17 | DEX0305_130 |
|    | DEX0305_22 | DEX0155_18      | DEX0305_131 |
| 40 | DEX0305_23 | DEX0155_19      |             |
|    | DEX0305_24 | DEX0155_20      | DEX0305_132 |
|    | DEX0305_25 | DEX0155_21      | DEX0305_133 |
|    | DEX0305_26 | DEX0155_22      | DEX0305_134 |
|    | DEX0305_27 | DEX0155_23      | DEX0305_135 |
| 45 | DEX0305_28 | flex DEX0155_23 |             |
|    | DEX0305_29 | DEX0155_24      | DEX0305_136 |

-110-

|    |            |                 |             |
|----|------------|-----------------|-------------|
|    | DEX0305_30 | DEX0155_25      | DEX0305_137 |
|    | DEX0305_31 | DEX0155_26      | DEX0305_138 |
|    | DEX0305_32 | DEX0155_27      | DEX0305_139 |
|    | DEX0305_33 | flex DEX0155_27 |             |
| 5  | DEX0305_34 | DEX0155_28      | DEX0305_140 |
|    | DEX0305_35 | DEX0155_29      | DEX0305_141 |
|    | DEX0305_36 | DEX0155_30      | DEX0305_142 |
|    | DEX0305_37 | flex DEX0155_30 | DEX0305_143 |
|    | DEX0305_38 | DEX0155_31      | DEX0305_144 |
| 10 | DEX0305_39 | DEX0155_32      | DEX0305_145 |
|    | DEX0305_40 | DEX0155_33      |             |
|    | DEX0305_41 | DEX0155_34      | DEX0305_146 |
|    | DEX0305_42 | DEX0155_35      |             |
|    | DEX0305_43 | DEX0155_36      | DEX0305_147 |
| 15 | DEX0305_44 | DEX0155_37      |             |
|    | DEX0305_45 | DEX0155_38      | DEX0305_148 |
|    | DEX0305_46 | DEX0155_39      | DEX0305_149 |
|    | DEX0305_47 | DEX0155_40      | DEX0305_150 |
|    | DEX0305_48 | flex DEX0155_40 | DEX0305_151 |
| 20 | DEX0305_49 | DEX0155_41      | DEX0305_152 |
|    | DEX0305_50 | DEX0155_42      | DEX0305_153 |
|    | DEX0305_51 | DEX0155_43      | DEX0305_154 |
|    | DEX0305_52 | DEX0155_44      | DEX0305_155 |
|    | DEX0305_53 | DEX0155_45      | DEX0305_156 |
| 25 | DEX0305_54 | DEX0155_46      | DEX0305_157 |
|    | DEX0305_55 | DEX0155_47      | DEX0305_158 |
|    | DEX0305_56 | flex DEX0155_47 |             |
|    | DEX0305_57 | DEX0155_48      | DEX0305_159 |
|    | DEX0305_58 | DEX0155_49      | DEX0305_160 |
| 30 | DEX0305_59 | DEX0155_50      | DEX0305_161 |
|    | DEX0305_60 | DEX0155_51      | DEX0305_162 |
|    | DEX0305_61 | DEX0155_52      | DEX0305_163 |
|    | DEX0305_62 | flex DEX0155_52 |             |
|    | DEX0305_63 | DEX0155_53      | DEX0305_164 |
| 35 | DEX0305_64 | DEX0155_54      | DEX0305_165 |
|    | DEX0305_65 | DEX0155_55      | DEX0305_166 |
|    | DEX0305_66 | DEX0155_56      | DEX0305_167 |
|    | DEX0305_67 | DEX0155_57      | DEX0305_168 |
|    | DEX0305_68 | DEX0155_58      | DEX0305_169 |
| 40 | DEX0305_69 | DEX0155_59      | DEX0305_170 |
|    | DEX0305_70 | DEX0155_60      | DEX0305_171 |
|    | DEX0305_71 | DEX0155_61      |             |
|    | DEX0305_72 | DEX0155_62      | DEX0305_172 |
|    | DEX0305_73 | DEX0155_63      | DEX0305_173 |
| 45 | DEX0305_74 | DEX0155_64      | DEX0305_174 |
|    | DEX0305_75 | flex DEX0155_64 |             |
|    | DEX0305_76 | DEX0155_65      | DEX0305_175 |
|    | DEX0305_77 | flex DEX0155_65 |             |
|    | DEX0305_78 | DEX0155_66      | DEX0305_176 |
| 50 | DEX0305_79 | DEX0155_67      | DEX0305_177 |
|    | DEX0305_80 | DEX0155_68      | DEX0305_178 |
|    | DEX0305_81 | DEX0155_69      | DEX0305_179 |
|    | DEX0305_82 | DEX0155_70      | DEX0305_180 |
|    | DEX0305_83 | DEX0155_71      | DEX0305_181 |
| 55 | DEX0305_84 | DEX0155_72      | DEX0305_182 |
|    | DEX0305_85 | DEX0155_73      | DEX0305_183 |
|    | DEX0305_86 | DEX0155_74      | DEX0305_184 |
|    | DEX0305_87 | DEX0155_75      | DEX0305_185 |
|    | DEX0305_88 | DEX0155_76      | DEX0305_186 |
| 60 | DEX0305_89 | flex DEX0155_76 | DEX0305_187 |
|    | DEX0305_90 | DEX0155_77      | DEX0305_188 |
|    | DEX0305_91 | DEX0155_78      | DEX0305_189 |

-111-

|    |             |                 |             |
|----|-------------|-----------------|-------------|
|    | DEX0305_92  | DEX0155_79      | DEX0305_190 |
|    | DEX0305_93  | DEX0155_80      | DEX0305_191 |
|    | DEX0305_94  | DEX0155_81      | DEX0305_192 |
|    | DEX0305_95  | flex DEX0155_81 |             |
| 5  | DEX0305_96  | DEX0155_82      | DEX0305_193 |
|    | DEX0305_97  | DEX0155_83      | DEX0305_194 |
|    | DEX0305_98  | DEX0155_84      | DEX0305_195 |
|    | DEX0305_99  | DEX0155_85      | DEX0305_196 |
|    | DEX0305_100 | DEX0155_86      | DEX0305_197 |
| 10 | DEX0305_101 | DEX0155_87      | DEX0305_198 |
|    | DEX0305_102 | DEX0155_88      | DEX0305_199 |
|    | DEX0305_103 | DEX0155_89      | DEX0305_200 |
|    | DEX0305_104 | DEX0155_90      |             |
|    | DEX0305_105 | flex DEX0155_90 |             |
| 15 | DEX0305_106 | DEX0155_94      | DEX0305_201 |
|    | DEX0305_107 | DEX0155_95      | DEX0305_202 |
|    | DEX0305_108 | DEX0155_96      | DEX0305_203 |
|    | DEX0305_109 | DEX0155_97      | DEX0305_204 |
|    | DEX0305_110 | DEX0155_98      | DEX0305_205 |
| 20 | DEX0305_111 | DEX0155_99      | DEX0305_206 |
|    | DEX0305_112 | DEX0155_100     | DEX0305_207 |
|    | DEX0305_113 | DEX0155_101     | DEX0305_208 |
|    | DEX0305_114 | DEX0155_102     | DEX0305_209 |
|    | DEX0305_115 | DEX0155_103     | DEX0305_210 |

**Example 1b: ATCC Deposit Information**

The table below summarizes the information corresponding to each BSG depicted in provisional application Serial No. 60/268,999, filed February 15, 2001, which is herein incorporated by reference in its entirety and which is referred to as DEX0155.

30 The cDNAs of the BSGs were deposited on the date listed in the column entitled ATCC Deposit Date. Each clone was cloned with vector PCR2.1 (Invitrogen, San Diego, CA). The "Contig Length" is the number of nucleotides in the contig identified by Contig ID and DEX0155 ID #. The "CloneSeq Length" is the number of nucleotides in the clone with "Clone ID" number and deposited with the ATCC.

35 The deposited material in the sample assigned ATCC Deposit Number in the table for any cDNA clone also contains one or more additional plasmids, each having a cDNA different from a given clone. Thus, deposits sharing the same ATCC number contain at least a plasmid for each "Clone ID" identified in the table. Typically, each ATCC deposit contains a mixture of approximately equal amounts by weight of about

40 fifty plasmids, each containing a different cDNA clone. The ATCC Deposit Number for ATCC breast pool 1 is PTA3060; the ATCC Deposit Number for ATCC breast pool 2 is PTA3061; the ATCC Deposit Number for ATCC breast pool 3 is PTA3062; and the ATCC Deposit Number for ATCC breast pool 4 is PTA3063.

The bioassays used were:

-112-

Psmam001.dc: This library consists of subtracted clones a pool of breast ductal cancer tissues, stage I, II, and II (two samples for each stage) versus cDNA from a pool of normal human tissues (spleen, pancreas, small intestine, heart, kidney, and liver).

Psmam002.dc: This library consists of subtracted cDNA clones from a pool of breast  
5 ductal cancer tissues, stage I, II, and II (two samples for each stage) versus a pool of normal human breast.

Psmam003.lc: This library consists of subtracted cDNA clones from a pool of breast lobular cancer tissues, three samples stage II versus a pool of other cancers (stomach, lung, and colon).

10 Psmam004.dc: This library consists of subtracted cDNA clones from a pool of breast ductal cancer tissues, three samples stage I, versus a pool of other cancers (stomach, lung, and colon).

Psmam005.dc: This library consists of subtracted cDNA clones from a pool of breast ductal cancer tissues, three samples stage I, versus a pool of other female cancers  
15 (uterus, cervix, endometrium, ovary).

Psmam006.dc: This library consists of subtracted cDNA clones from a pool of breast ductal cancer tissues, three samples stage I, versus a pool of normal human breast.

Psmam007.lc: This library consists of subtracted cDNA clones from a pool of breast lobular cancer tissues, three samples stage II versus a pool of normal human breast.

20 Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in the Table below. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to clone id, e.g., 601537248F1.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using  
25 an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance with 33P-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g. Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue  
30 (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated in 1.5% agar plates (containing the appropriate selection agent, e.g. ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony

-113-

screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two printers of 17-20 nucleotides derived from both ends of the  
5 DEX0155 ID NO:X (i.e., within the region of DEX0155 ID NO:X bounded by the 5' NT  
and the 3'NT of the clone defined in the table below) are synthesized and used to amplify  
the desired cDNA using the deposited cDNA plasmid as a template. The polymerase  
chain reaction is carried out under routine conditions, for instance, in 25ul of reaction  
mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-  
10 5mM MgCl<sub>2</sub>, 0.01% (w/v) gelatin, 20uM each of dATP, dCTP, dGTP, dTTP, 25 pmol of  
each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at  
94°C for 1 minute; annealing at 55°C for 1 minute; elongation at 72°C for 1 minute) are  
performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product  
is analyzed by agarose gel electrophoresis and the DNA band with expected molecular  
15 weight is excised and purified. The PCR product is verified to be the selected sequence  
by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding  
portions of a gene which may not be present in the deposited clone. These methods  
include but are not limited to, filter probing, clone enrichment using specific probes, and  
20 protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the  
art. For instance, a method similar or identical to 5' RACE is available for generating the  
missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids  
Res.* 21(7); 1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of  
25 RNA presumably containing full-length gene RNA transcripts. A primer set containing a  
primer specific to the ligated RNA oligonucleotide and a primer specific to a known  
sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-  
length gene. This amplified product may then be sequenced and used to generate the full  
length gene.

30 This above method starts with total RNA isolated from the desired source,  
although poly-A+RNA can be used. The RNA preparation can then be treated with  
phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA  
which may interfere with the later RNA ligase step. The phosphatase should then be  
inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove

the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

- This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

| Seq ID | ContigID | CntgLn<br>gth | cloneID     | clnSLe<br>Length<br>Lg | Deposit Date      | ATCC pool         | Bio Assay   |
|--------|----------|---------------|-------------|------------------------|-------------------|-------------------|-------------|
| 1      | 6.1      | 959           | 601537248F1 | 475                    | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|        | 6.1      | 959           | 601542790F1 | 531                    | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|        | 6.1      | 959           | 601525159F1 | 959                    | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
| 2      | 6.1009   | 1079          | 601540888F1 | 1174                   |                   |                   | PSmam007.lc |
|        | 6.1009   | 1079          | 601540788F1 | 1020                   | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 3      | 6.1011   | 1782          | 601543772F1 | 798                    | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|        | 6.1011   | 1782          | 601536151F1 | 1876                   | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 4      | 6.1072   | 1023          | 601540655F1 | 983                    | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|        | 6.1072   | 1023          | 601540754F1 | 1055                   | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 5      | 6.108    | 807           | 601528950F1 | 807                    | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 6      | 6.1089   | 354           | 601541688F1 | 538                    |                   |                   | PSmam007.lc |
|        | 6.1089   | 354           | 601536741F1 | 952                    |                   |                   | PSmam007.lc |
| 7      | 6.1103   | 540           | 601539887F1 | 948                    |                   |                   | PSmam007.lc |
|        | 6.1103   | 540           | 601538875F1 | 609                    | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 8      | 6.111    | 292           | 601602510F1 | 292                    |                   |                   | PSmam001.dc |
| 9      | 6.1122   | 645           | 601537921F1 | 713                    |                   |                   | PSmam007.lc |
|        | 6.1122   | 645           | 601540113F1 | 1057                   |                   |                   | PSmam007.lc |
| 10     | 6.1126   | 122           | 601540259F1 | 761                    |                   |                   | PSmam007.lc |
|        | 6.1126   | 122           | 601538183F1 | 665                    |                   |                   | PSmam007.lc |
| 11     | 6.113    | 785           | 601597950F1 | 785                    |                   |                   | PSmam001.dc |
| 12     | 6.114    | 863           | 601517518F1 | 863                    | February 15, 2001 | ATCC-breast pool2 | PSmam005.dc |
| 13     | 6.1155   | 861           | 601543573F1 | 929                    | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|        | 6.1155   | 861           | 601542440F1 | 604                    | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 14     | 6.1208   | 1009          | 601538978F1 | 1074                   | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|        | 6.1208   | 1009          | 601531017F1 | 900                    | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 15     | 6.1243   | 357           | 601603586F1 | 676                    |                   |                   | PSmam001.dc |
|        | 6.1243   | 357           | 600952042F1 | 669                    |                   |                   | PSmam001.dc |
|        | 6.1243   | 357           | 600366720   | 502                    |                   |                   | PSmam001.dc |
| 16     | 6.125    | 896           | 601532490F1 | 896                    | February 15, 2001 | ATCC-breast pool1 | PSmam006.dc |
| 17     | 6.132    | 415           | 601526961F1 | 541                    |                   |                   | PSmam006.dc |
|        | 6.132    | 415           | 601526661F1 | 760                    |                   |                   | PSmam006.dc |
| 18     | 6.133    | 1049          | 601527177F1 | 1049                   |                   |                   | PSmam006.dc |
| 19     | 6.1466   | 893           | 601602565F1 | 972                    |                   |                   | PSmam001.dc |
|        | 6.1466   | 893           | 601604084F1 | 931                    |                   |                   | PSmam001.dc |
| 20     | 6.1557   | 694           | 601522418F1 | 562                    | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
|        | 6.1557   | 694           | 601522318F1 | 730                    |                   |                   | PSmam005.dc |
| 21     | 6.1679   | 542           | 601017282F1 | 400                    |                   |                   | PSmam003.lb |
|        | 6.1679   | 542           | 601526928F1 | 714                    | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |



|    |        |      |             |      |                   |                   |             |
|----|--------|------|-------------|------|-------------------|-------------------|-------------|
| 22 | 6.168  | 923  | 601539979F1 | 923  |                   |                   | PSmam007.lc |
| 23 | 6.1687 | 326  | 601525041F1 | 658  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
|    | 6.1687 | 326  | 601525133F1 | 730  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
| 24 | 6.169  | 933  | 601525912F1 | 933  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
| 25 | 6.1697 | 603  | 601514667F1 | 997  | February 15, 2001 | ATCC-breast pool2 | PSmam005.dc |
|    | 6.1697 | 603  | 601521028F1 | 635  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 26 | 6.1717 | 136  | 601514659F1 | 385  | February 15, 2001 | ATCC-breast pool2 | PSmam005.dc |
|    | 6.1717 | 136  | 601522657F1 | 664  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 27 | 6.172  | 979  | 601526109F1 | 979  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
| 28 | 6.1795 | 933  | 601529339F1 | 1032 | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.1795 | 933  | 601524119F1 | 849  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
|    | 6.1795 | 933  | 601518117F1 | 1041 |                   |                   | PSmam005.dc |
| 29 | 6.1819 | 911  | 601540162F1 | 878  |                   |                   | PSmam007.lc |
|    | 6.1819 | 911  | 601535132F1 | 920  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.1819 | 911  | 601540145F1 | 941  |                   |                   | PSmam007.lc |
| 30 | 6.182  | 475  | 601541943F1 | 1583 | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.182  | 475  | 601597722F1 | 559  |                   |                   | PSmam001.dc |
|    | 6.182  | 475  | 601596465F1 | 693  |                   |                   | PSmam001.dc |
| 31 | 6.1917 | 709  | 601544053F1 | 655  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.1917 | 709  | 601540334F1 | 938  |                   |                   | PSmam007.lc |
|    | 6.1917 | 709  | 601542665F1 | 942  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 32 | 6.1935 | 722  | 601541707F1 | 820  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.1935 | 722  | 601536777F1 | 941  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.1935 | 722  | 601538649F1 | 805  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 33 | 6.198  | 954  | 601537453F1 | 954  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 34 | 6.1983 | 247  | 601543641F1 | 1123 | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.1983 | 247  | 601538684F1 | 701  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.1983 | 247  | 601535158F1 | 891  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
| 35 | 6.1992 | 528  | 601516523F1 | 703  |                   |                   | PSmam005.dc |
|    | 6.1992 | 528  | 601514351F1 | 696  | February 15, 2001 | ATCC-breast pool2 | PSmam005.dc |
|    | 6.1992 | 528  | 601524114F1 | 751  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 36 | 6.2    | 919  | 601535856F1 | 919  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 37 | 6.201  | 890  | 601536584F1 | 796  |                   |                   | PSmam007.lc |
|    | 6.201  | 890  | 601536284F1 | 761  |                   |                   | PSmam007.lc |
|    | 6.201  | 890  | 601535587F1 | 959  |                   |                   | PSmam007.lc |
| 38 | 6.2045 | 387  | 601523866F1 | 639  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
|    | 6.2045 | 387  | 601516660F1 | 832  |                   |                   | PSmam005.dc |
|    | 6.2045 | 387  | 601518993F1 | 841  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 39 | 6.207  | 728  | 601536544F1 | 728  |                   |                   | PSmam007.lc |
| 40 | 6.2092 | 502  | 601537436F1 | 724  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.2092 | 502  | 601544184F1 | 741  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.2092 | 502  | 601543801F1 | 887  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
| 41 | 6.21   | 344  | 601596828F1 | 344  |                   |                   | PSmam001.dc |
| 42 | 6.2135 | 645  | 601534191F1 | 772  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2135 | 645  | 601532470F1 | 646  | February 15, 2001 | ATCC-breast pool2 | PSmam006.dc |
|    | 6.2135 | 645  | 601530937F1 | 742  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 43 | 6.2158 | 173  | 601600462F1 | 694  |                   |                   | PSmam004.dc |
|    | 6.2158 | 173  | 601600603F1 | 638  |                   |                   | PSmam004.dc |
|    | 6.2158 | 173  | 601600341F1 | 778  |                   |                   | PSmam004.dc |
| 44 | 6.216  | 430  | 601601334F1 | 430  |                   |                   | PSmam004.dc |
| 45 | 6.2167 | 1314 | 601531861F1 | 409  |                   |                   | PSmam006.dc |
|    | 6.2167 | 1314 | 601525253F1 | 1360 | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
|    | 6.2167 | 1314 | 601528683F1 | 205  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 46 | 6.2194 | 418  | 601597228F1 | 466  |                   |                   | PSmam001.dc |
|    | 6.2194 | 418  | 601596539F1 | 579  |                   |                   | PSmam001.dc |

|    |        |     |             |      |                   |                   |             |
|----|--------|-----|-------------|------|-------------------|-------------------|-------------|
|    | 6.2194 | 418 | 601597336F1 | 495  |                   |                   | PSmam001.dc |
| 47 | 6.2236 | 672 | 601518414F1 | 735  | February 15, 2001 | ATCC-breast pool2 | PSmam005.dc |
|    | 6.2236 | 672 | 601523707F1 | 771  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
|    | 6.2236 | 672 | 601521237F1 | 963  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 48 | 6.2324 | 687 | 601519403F1 | 704  |                   |                   | PSmam005.dc |
|    | 6.2324 | 687 | 601598936F1 | 555  |                   |                   | PSmam001.dc |
|    | 6.2324 | 687 | 601514372F1 | 885  | February 15, 2001 | ATCC-breast pool2 | PSmam005.dc |
|    | 6.2324 | 687 | 601519248F1 | 763  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 49 | 6.2342 | 219 | 601525103F1 | 558  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
|    | 6.2342 | 219 | 601528288F1 | 730  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2342 | 219 | 601528520F1 | 730  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2342 | 219 | 601530055F1 | 657  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 50 | 6.2416 | 470 | 601596393F1 | 462  |                   |                   | PSmam001.dc |
|    | 6.2416 | 470 | 601016531F1 | 1066 |                   |                   | PSmam003.lb |
|    | 6.2416 | 470 | 601019392F1 | 463  |                   |                   | PSmam004.dc |
|    | 6.2416 | 470 | 601597524F1 | 518  |                   |                   | PSmam001.dc |
| 51 | 6.242  | 934 | 601529021F1 | 934  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 52 | 6.247  | 959 | 601603531F1 | 959  |                   |                   | PSmam001.dc |
| 53 | 6.253  | 738 | 601527386F1 | 738  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 54 | 6.2564 | 713 | 601537450F1 | 681  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.2564 | 713 | 601534611F1 | 764  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2564 | 713 | 601526037F1 | 787  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
|    | 6.2564 | 713 | 601526069F1 | 768  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
| 55 | 6.276  | 601 | 601534954F1 | 601  |                   |                   | PSmam006.dc |
| 56 | 6.277  | 323 | 601521708F1 | 323  |                   |                   | PSmam005.dc |
| 57 | 6.281  | 671 | 601529751F1 | 671  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 58 | 6.2816 | 488 | 601530322F1 | 604  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2816 | 488 | 601526430F1 | 560  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
|    | 6.2816 | 488 | 601531659F1 | 711  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2816 | 488 | 601534668F1 | 591  |                   |                   | PSmam006.dc |
| 59 | 6.2847 | 487 | 601530032F1 | 645  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2847 | 487 | 601533983F1 | 789  |                   |                   | PSmam006.dc |
|    | 6.2847 | 487 | 601528311F1 | 644  |                   |                   | PSmam006.dc |
|    | 6.2847 | 487 | 601533391F1 | 531  |                   |                   | PSmam006.dc |
| 60 | 6.2908 | 667 | 601521173F1 | 904  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
|    | 6.2908 | 667 | 601520675F1 | 963  |                   |                   | PSmam005.dc |
|    | 6.2908 | 667 | 601519131F1 | 815  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
|    | 6.2908 | 667 | 601521583F1 | 718  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 61 | 6.2952 | 802 | 601532320F1 | 615  | February 15, 2001 | ATCC-breast pool1 | PSmam006.dc |
|    | 6.2952 | 802 | 601533462F1 | 823  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.2952 | 802 | 601532519F1 | 752  | February 15, 2001 | ATCC-breast pool1 | PSmam006.dc |
|    | 6.2952 | 802 | 601533719F1 | 927  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 62 | 6.2992 | 918 | 600955661F1 | 910  |                   |                   | PSmam002.dc |
|    | 6.2992 | 918 | 601603160F1 | 721  |                   |                   | PSmam001.dc |
|    | 6.2992 | 918 | 601544238F1 | 627  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.2992 | 918 | 601602784F1 | 702  |                   |                   | PSmam001.dc |
| 63 | 6.302  | 716 | 601600667F1 | 716  |                   |                   | PSmam004.dc |
| 64 | 6.303  | 348 | 600371641   | 348  |                   |                   | PSmam004.dc |
| 65 | 6.3078 | 985 | 601539380F1 | 734  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.3078 | 985 | 601545391F1 | 723  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.3078 | 985 | 601541726F1 | 973  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.3078 | 985 | 601539747F1 | 961  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 66 | 6.313  | 903 | 601531005F1 | 903  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 67 | 6.3226 | 669 | 601519607F1 | 727  |                   |                   | PSmam005.dc |
|    | 6.3226 | 669 | 601521024F1 | 731  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |

-117-

|    |        |      |             |      |                   |                   |             |
|----|--------|------|-------------|------|-------------------|-------------------|-------------|
|    | 6.3226 | 669  | 601523328F1 | 761  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
|    | 6.3226 | 669  | 601520103F1 | 724  | February 15, 2001 | ATCC-breast pool3 | PSmam005.dc |
| 68 | 6.326  | 928  | 601515770F1 | 928  |                   |                   | PSmam005.dc |
| 69 | 6.33   | 928  | 601515523F1 | 928  |                   |                   | PSmam005.dc |
| 70 | 6.3305 | 837  | 601532496F1 | 815  | February 15, 2001 | ATCC-breast pool2 | PSmam006.dc |
|    | 6.3305 | 837  | 601595830F1 | 627  |                   |                   | PSmam001.dc |
|    | 6.3305 | 837  | 601534496F1 | 699  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.3305 | 837  | 601596661F1 | 526  |                   |                   | PSmam001.dc |
| 71 | 6.348  | 254  | 600945579F1 | 254  |                   |                   | PSmam003.lb |
| 72 | 6.349  | 511  | 601602521F1 | 511  |                   |                   | PSmam001.dc |
| 73 | 6.356  | 773  | 601603632F1 | 773  |                   |                   | PSmam001.dc |
| 74 | 6.362  | 881  | 601524722F1 | 881  |                   |                   | PSmam006.dc |
| 75 | 6.366  | 815  | 601527869F1 | 815  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 76 | 6.378  | 718  | 601544680F1 | 720  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.378  | 718  | 601544887F1 | 848  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.378  | 718  | 601544772F1 | 718  |                   |                   | PSmam007.lc |
|    | 6.378  | 718  | 601544695F1 | 733  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
| 77 | 6.3861 | 976  | 601608403F1 | 743  |                   |                   | PSmam002.dc |
|    | 6.3861 | 976  | 601608269F1 | 714  |                   |                   | PSmam002.dc |
|    | 6.3861 | 976  | 601607621F1 | 697  |                   |                   | PSmam002.dc |
|    | 6.3861 | 976  | 601603679F1 | 624  |                   |                   | PSmam001.dc |
| 78 | 6.39   | 823  | 601602589F1 | 823  |                   |                   | PSmam001.dc |
| 79 | 6.396  | 671  | 601530872F1 | 671  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 80 | 6.413  | 533  | 601017869F1 | 533  |                   |                   | PSmam003.lb |
| 81 | 6.417  | 506  | 601607908F1 | 506  |                   |                   | PSmam002.dc |
| 82 | 6.424  | 1494 | 601539527F1 | 1494 | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 83 | 6.436  | 813  | 601595862F1 | 813  |                   |                   | PSmam001.dc |
| 84 | 6.449  | 1578 | 601542644F1 | 1578 | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 85 | 6.452  | 779  | 601540769F1 | 779  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
| 86 | 6.461  | 812  | 601598914F1 | 785  |                   |                   | PSmam001.dc |
| 87 | 6.653  | 377  | 601527708F1 | 599  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.653  | 377  | 601526534F1 | 794  | February 15, 2001 | ATCC-breast pool3 | PSmam006.dc |
| 88 | 6.714  | 1527 | 600370346   | 345  |                   |                   | PSmam003.lb |
|    | 6.714  | 1527 | 601538741F1 | 1472 | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.714  | 1527 | 600370396   | 472  |                   |                   | PSmam003.lb |
| 89 | 6.782  | 532  | 601535238F1 | 729  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.782  | 532  | 601542856F1 | 668  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
| 90 | 6.81   | 499  | 601542724F1 | 686  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
|    | 6.81   | 499  | 601542984F1 | 752  | February 15, 2001 | ATCC-breast pool2 | PSmam007.lc |
| 91 | 6.848  | 544  | 601528693F1 | 611  |                   |                   | PSmam006.dc |
|    | 6.848  | 544  | 601528851F1 | 583  |                   |                   | PSmam006.dc |
| 92 | 6.853  | 212  | 601528335F1 | 671  |                   |                   | PSmam006.dc |
|    | 6.853  | 212  | 601601048F1 | 616  |                   |                   | PSmam004.dc |
| 93 | 6.862  | 658  | 601517588F1 | 853  | February 15, 2001 | ATCC-breast pool2 | PSmam005.dc |
|    | 6.862  | 658  | 601598143F1 | 729  |                   |                   | PSmam001.dc |
| 94 | 6.872  | 226  | 601529142F1 | 329  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
|    | 6.872  | 226  | 601529976F1 | 472  | February 15, 2001 | ATCC-breast pool4 | PSmam006.dc |
| 95 | 6.874  | 430  | 601601839F1 | 640  |                   |                   | PSmam004.dc |
|    | 6.874  | 430  | 601516279F1 | 502  |                   |                   | PSmam005.dc |
| 96 | 6.944  | 813  | 601600731F1 | 1105 |                   |                   | PSmam004.dc |
|    | 6.944  | 813  | 601600831F1 | 837  |                   |                   | PSmam004.dc |
| 97 | 6.958  | 444  | 601044611F1 | 833  |                   |                   | PSmam002.dc |
|    | 6.958  | 444  | 601608144F1 | 682  |                   |                   | PSmam002.dc |
| 98 | 6.981  | 273  | 601042054F1 | 639  |                   |                   | PSmam004.dc |
|    | 6.981  | 273  | 601599360F1 | 838  |                   |                   | PSmam004.dc |

-118-

|    |       |      |             |      |                   |                   |             |
|----|-------|------|-------------|------|-------------------|-------------------|-------------|
|    | 6.981 | 273  | 601041167F1 | 790  |                   |                   | PSmam004.dc |
|    | 6.981 | 273  | 601042165F1 | 667  |                   |                   | PSmam004.dc |
| 99 | 6.99  | 1397 | 601542353F1 | 1464 | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |
|    | 6.99  | 1397 | 601541284F1 | 764  | February 15, 2001 | ATCC-breast pool1 | PSmam007.lc |

### Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation  
 5 detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The  
 method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5'  
 reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity  
 of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected  
 by the laser detector of the Model 7700 Sequence Detection System (PE Applied  
 10 Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to  
 standardize the amount of sample RNA added to the reaction and normalize for Reverse  
 Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate  
 dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this  
 endogenous control. To calculate relative quantitation between all the samples studied,  
 15 the target RNA levels for one sample were used as the basis for comparative results  
 (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard  
 curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence  
 Detection System).

The tissue distribution and the level of the target gene are evaluated for every  
 20 sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer  
 tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently,  
 first strand cDNA is prepared with reverse transcriptase and the polymerase chain  
 reaction is done using primers and Taqman probes specific to each target gene. The  
 results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute  
 25 numbers are relative levels of expression of the target gene in a particular tissue  
 compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of  
 expression of the BSNA versus normal tissues and other cancer tissues can then be  
 determined. All the values are compared to a normal tissue (calibrator). These RNA  
 30 samples are commercially available pools, originated by pooling samples of a particular  
 tissue from different individuals.

-119-

The relative levels of expression of the BSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to a normal tissue (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for  
5 that same tissue from the same individual.

In the analysis of matching samples, BSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal  
10 adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 115 being diagnostic  
15 markers for cancer.

#### **Example 3: Protein Expression**

The BSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the BSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the BSNA coding sequence, codons for two amino acids, Met-Ala, flanking  
20 the NH<sub>2</sub>-terminus of the coding sequence of BSNA, and six histidines, flanking the COOH-terminus of the coding sequence of BSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is  
25 confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of BSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column  
30 volumes of wash buffer. BSP was eluted stepwise with various concentration imidazole buffers.

#### **Example 4: Protein Fusions**

-120-

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

**Example 5: Production of an Antibody from a Polypeptide**

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method,

-121-

protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such

- 5 antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

- 10 The predicted antigenicity for the amino acid sequences is as follows:

| DEX ID      | ANTIGENICITY  | TRANSMEMBRANE                   | PTM  | SIGNAL PEPTIDE                        |
|-------------|---|---------------------------------|--|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length                              | Predicted<br>Helix,<br>Topology | PTM  | Position,<br>Max Score,<br>Mean Score |
| DEX0305_116 | 10-46, 1.06, 37   |                                 | Ck2_Phospho_Site 33-36; Myristyl 45-50; Tyr_Phospho_Site 53-59;  |                                       |
| DEX0305_117 | 350-381, 1.06, 32<br>119-137, 1.04, 19<br>290-324, 1.03, 35 |                                 | Amidation 263-266;<br>Asn_Glycosylation 90-93; 95-98;<br>Camp_Phospho_Site 83-86; 144-147; 265-268;<br>Ck2_Phospho_Site 5-8; 14-17; 68-71; 72-75; 205-208; 328-331;<br>Myristyl 290-295;<br>Pkc_Phospho_Site 142-144; 174-176; 187-189; 268-270; 352-354; 405-407;<br>Tyr_Phospho_Site 15-22; 104-112; 189-196;<br>Zinc_Finger_C2h2 260-280; 288-308; 316-336; 344-364; 372-392; |                                       |
| DEX0305_118 | 41-52, 1.09, 12   |                                 | Camp_Phospho_Site 99-102;<br>Cytochrome_C 117-122; Myristyl 65-70;<br>Pkc_Phospho_Site 97-99; 102-104; 112-114;<br>Prokar_Lipoprotei   | 17, .964, .707                        |

-122-

| DEX ID      | ANTIGENICITY                   | TRANSMEMBRANE                   | PTM  | SIGNAL PEPTIDE                        |
|-------------|--------------------------------|---------------------------------|--|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length | Predicted<br>Helix,<br>Topology | PTM  | Position,<br>Max Score,<br>Mean Score |
| DEX0305_119 |                                |                                 | n 60-70;<br>Amidation 95-98;<br>Ck2_Phospho_Site 16-19;<br>Pkc_Phospho_Site 16-18;20-22;27-29;33-35;107-109;108-110;119-121; |                                       |
| DEX0305_120 |                                |                                 | Camp_Phospho_Site 12-15;   |                                       |
| DEX0305_121 | 58-70,1.07,13<br>9-22,1.06,14  |                                 | Asn_Glycosylation 51-54;<br>Ck2_Phospho_Site 37-40; Myristyl 33-38;<br>Pkc_Phospho_Site 10-12;91-93;                         |                                       |
| DEX0305_122 |                                |                                 | Ck2_Phospho_Site 3-6;  |                                       |
| DEX0305_123 | 24-42,1.31,19                  |                                 | Amidation 26-29;<br>Asn_Glycosylation 34-37;<br>Camp_Phospho_Site 28-31; Myristyl 26-31;<br>Pkc_Phospho_Site 30-32;          |                                       |
| DEX0305_124 | 16-26,1.07,11                  |                                 | Ck2_Phospho_Site 34-37;<br>Pkc_Phospho_Site 19-21;   |                                       |
| DEX0305_125 | 20-30,1.06,11                  |                                 | Ck2_Phospho_Site 20-23;57-60;<br>Myristyl 66-71;   |                                       |
| DEX0305_126 |                                |                                 | Asn_Glycosylation 14-17;<br>Camp_Phospho_Site 5-8;<br>Ck2_Phospho_Site 21-24;<br>Pkc_Phospho_Site 16-18;25-27;               |                                       |
| DEX0305_128 |                                |                                 | Ck2_Phospho_Site 26-29;<br>Pkc_Phospho_Site 12-14;   | 18,.967,.859                          |
| DEX0305_129 |                                |                                 | Asn_Glycosylation 29-32;<br>Pkc_Phospho_Site 15-17;42-44;49-51;  |                                       |
| DEX0305_130 | 131-155,1,25                   |                                 | Asn_Glycosylation 112-115;<br>Ck2_Phospho_Site 10-13;37-40;149-  |                                       |



-123-

| DEX ID      | ANTIGENICITY                           | TRANSMEMBRANE                   | PTM  | SIGNAL PEPTIDE                        |
|-------------|--|---------------------------------|--|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length         | Predicted<br>Helix,<br>Topology | PTM  | Position,<br>Max Score,<br>Mean Score |
|             |  |                                 | 152;<br>Pkc_Phospho_Site<br>37-39;<br>Tyr_Phospho_Site<br>19-27;   |                                       |
| DEX0305_132 |  |                                 | Camp_Phospho_Site<br>44-47;<br>Ck2_Phospho_Site<br>47-50;<br>Pkc_Phospho_Site<br>42-44;49-51;  |                                       |
| DEX0305_133 | 10-<br>28,1.09,19                      |                                 | Ck2_Phospho_Site<br>12-15;<br>Pkc_Phospho_Site<br>16-18;39-41;   |                                       |
| DEX0305_135 | 56-<br>96,1.12,41                      | 1,0106-128i                     | Asn_Glycosylation<br>58-61;131-134;<br>Camp_Phospho_Site<br>70-73;<br>Ck2_Phospho_Site<br>44-47;74-77;<br>Glycosaminoglycan<br>75-78; Myristyl<br>64-69;<br>Pkc_Phospho_Site<br>68-70;87-89; |                                       |
| DEX0305_136 |  |                                 | Camp_Phospho_Site<br>20-23; Myristyl<br>28-33;<br>Pkc_Phospho_Site<br>4-6;10-12;11-<br>13;19-21;   |                                       |
| DEX0305_137 |  |                                 | Pkc_Phospho_Site<br>16-18;   |                                       |
| DEX0305_138 |  |                                 | Ck2_Phospho_Site<br>44-47;   |                                       |
| DEX0305_139 | 41-<br>55,1.18,15                      |                                 | Asn_Glycosylation<br>41-44; Myristyl<br>30-35;56-61;<br>Pkc_Phospho_Site<br>11-13;80-82;   |                                       |
| DEX0305_140 |  | 1,i113-35o                      | Ck2_Phospho_Site<br>60-63; Myristyl<br>49-54;<br>Pkc_Phospho_Site<br>50-52;60-62;  |                                       |
| DEX0305_141 | 59-<br>69,1.14,11<br>30-<br>46,1.08,17 |                                 | Ck2_Phospho_Site<br>29-32; Myristyl<br>51-56;<br>Pkc_Phospho_Site<br>59-61;<br>Prokar_Lipoprotein<br>50-60;  |                                       |
| DEX0305_143 | 39-51,1,13                             |                                 | Ck2_Phospho_Site<br>39-42; Myristyl<br>86-91;  |                                       |

-124-

| DEX ID      | ANTIGENICITY                   | TRANSMEMBRANE                   | PTM  | SIGNAL PEPTIDE                        |
|-------------|--------------------------------|---------------------------------|--|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length | Predicted<br>Helix,<br>Topology | PTM  | Position,<br>Max Score,<br>Mean Score |
|             |                                |                                 | Pkc_Phospho_Site<br>5-7;26-28;   |                                       |
| DEX0305_144 | 88-<br>193,1.09,10<br>6        |                                 | Amidation 78-<br>81;174-177;<br>Asn_Glycosylation<br>109-112;<br>Camp_Phospho_Site<br>150-153;151-154;<br>Ck2_Phospho_Site<br>2-5;57-60;91-<br>94;130-133;153-<br>156;154-157;168-<br>171;<br>Pkc_Phospho_Site<br>44-46;57-59;67-<br>69;110-112;130-<br>132;162-164;168-<br>170;   |                                       |
| DEX0305_145 |                                |                                 | Myristyl 67-72;<br>Pkc_Phospho_Site<br>18-20;31-33;<br>Tyr_Phospho_Site<br>66-73;  |                                       |
| DEX0305_146 | 38-<br>51,1.19,14              |                                 | Asn_Glycosylation<br>39-42;<br>Pkc_Phospho_Site<br>29-31;41-43;  |                                       |
| DEX0305_147 | 29-<br>42,1.17,14              |                                 | Myristyl 30-35;  |                                       |
| DEX0305_148 |                                |                                 | Amidation 3-6;   |                                       |
| DEX0305_149 |                                | 1,038-55i                       | Ck2_Phospho_Site<br>28-31; Myristyl<br>10-15;78-83;  |                                       |
| DEX0305_150 |                                |                                 | Asn_Glycosylation<br>10-13;<br>Pkc_Phospho_Site<br>82-84;  |                                       |
| DEX0305_151 | 95-<br>126,1.11,32             |                                 | Camp_Phospho_Site<br>123-126;<br>Ck2_Phospho_Site<br>27-30;33-36;58-<br>61;93-96;165-<br>168;171-174;213-<br>216;225-228;226-<br>229; Myristyl 9-<br>14;17-22;23-<br>28;39-44;84-<br>89;155-160;161-<br>166;166-171;177-<br>182;247-252;<br>Pkc_Phospho_Site<br>13-15;69-71;151-<br>153;206-208;213-<br>215;235-237;251- |                                       |

-125-

| DEX ID      | ANTIGENICITY                   | TRANSMEMBRANE                   | PTM   | SIGNAL PEPTIDE                        |
|-------------|--------------------------------|---------------------------------|---|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length | Predicted<br>Helix,<br>Topology | PTM   | Position,<br>Max Score,<br>Mean Score |
| DEX0305_152 |                                |                                 | 253;<br>Myristyl 7-12;30-35;48-53;51-56;56-61;57-62;61-66;72-77;90-95;<br>Pkc_Phospho_Site 91-93;   | 19,.995,.853                          |
| DEX0305_153 |                                |                                 |   | 18,.98,.909                           |
| DEX0305_154 | 3-60,1.09,58                   |                                 | Asn_Glycosylation 18-21;<br>Camp_Phospho_Site 54-57;<br>Ck2_Phospho_Site 38-41;40-43;<br>Pkc_Phospho_Site 12-14;40-42;                        |                                       |
| DEX0305_155 |                                |                                 | Asn_Glycosylation 9-12;<br>Camp_Phospho_Site 7-10;<br>Ck2_Phospho_Site 10-13;<br>Pkc_Phospho_Site 45-47;                                      |                                       |
| DEX0305_156 |                                | 2,i17-36o56-78i                 | Myristyl 10-15;27-32;75-80;<br>Pkc_Phospho_Site 66-68;  |                                       |
| DEX0305_157 |                                |                                 | Ck2_Phospho_Site 17-20;<br>Prokar_Lipoprotein 8-18;   |                                       |
| DEX0305_158 |                                | 1,o39-61i                       | Pkc_Phospho_Site 20-22;   |                                       |
| DEX0305_159 |                                |                                 | Ck2_Phospho_Site 25-28;50-53;<br>Pkc_Phospho_Site 24-26;43-45;50-52;  |                                       |
| DEX0305_160 |                                |                                 | Asn_Glycosylation 90-93;<br>Camp_Phospho_Site 33-36;<br>Ck2_Phospho_Site 25-28;92-95;<br>Myristyl 57-62;79-84;<br>Pkc_Phospho_Site 6-8;43-45; | 19,.91,.704                           |
| DEX0305_161 |                                |                                 | Myristyl 8-13;  |                                       |
| DEX0305_162 |                                |                                 | Amidation 14-17;<br>Ck2_Phospho_Site 35-38;   |                                       |

-126-

| DEX ID      | ANTIGENICITY                   | TRANSMEMBRANE                                      | PTM   | SIGNAL PEPTIDE                        |
|-------------|--------------------------------|--|---|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length | Predicted<br>Helix,<br>Topology                    | PTM   | Position,<br>Max Score,<br>Mean Score |
|             |                                |  | Glycosaminoglycan<br>6-9; Myristyl 2-<br>7;9-14;  |                                       |
| DEX0305_163 |                                | 1,015-37i  | Myristyl 14-19;<br>Pkc_Phospho_Site<br>66-68;   |                                       |
| DEX0305_164 |                                | 1,010-32i  |   |                                       |
| DEX0305_165 |                                | 1,05-27i   |   |                                       |
| DEX0305_166 |                                |  | Ck2_Phospho_Site<br>22-25;<br>Pkc_Phospho_Site<br>22-24;  | 17,.938,.8<br>26                      |
| DEX0305_167 | 8-<br>17,1.03,10               |  |   |                                       |
| DEX0305_168 |                                | 5,17-29o44-<br>63i100-<br>122o127-<br>149i151-173o | Camp_Phospho_Site<br>81-84;<br>Leucine_Zipper 4-<br>25;122-143;140-<br>161;147-168;<br>Myristyl 202-207;<br>Pkc_Phospho_Site<br>80-82;96-98;208-<br>210;<br>Prokar_Lipoprotei<br>n 22-32;166-176; | 21,.998,.9<br>42                      |
| DEX0305_170 |                                |  | Myristyl 2-7;<br>Pkc_Phospho_Site<br>14-16;   |                                       |
| DEX0305_172 |                                |  | Myristyl 12-17;<br>Pkc_Phospho_Site<br>28-30;39-41;   | 22,.887,.5<br>85                      |
| DEX0305_173 |                                |  | Asn_Glycosylation<br>53-56;<br>Camp_Phospho_Site<br>40-43;<br>Ck2_Phospho_Site<br>29-32; Myristyl<br>12-17;47-52;<br>Pkc_Phospho_Site<br>16-18;   |                                       |
| DEX0305_174 |                                |  | Asn_Glycosylation<br>41-44;<br>Ck2_Phospho_Site<br>8-11; Myristyl<br>42-47;52-57;<br>Phosphopantethein<br>e 32-47;<br>Pkc_Phospho_Site<br>29-31;  |                                       |
| DEX0305_175 | 52-63,1,12                     |  | Pkc_Phospho_Site<br>52-54;  |                                       |
| DEX0305_176 |                                |  | Asn_Glycosylation<br>25-28;   |                                       |

-127-

| DEX ID      | ANTIGENICITY                   | TRANSMEMBRANE                   | PTM  | SIGNAL PEPTIDE                        |
|-------------|--------------------------------|---------------------------------|--|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length | Predicted<br>Helix,<br>Topology | PTM  | Position,<br>Max Score,<br>Mean Score |
|             |                                |                                 | Ck2_Phospho_Site<br>29-32;<br>Pkc_Phospho_Site<br>43-45;   |                                       |
| DEX0305_177 | 89-119,1,31                    |                                 | Myristyl 70-75;105-110;<br>Pkc_Phospho_Site<br>10-12;21-23;<br>Prokar_Lipoprotein 42-52; Rgd 92-94;                        | 50,.925,.649                          |
| DEX0305_178 |                                |                                 | Camp_Phospho_Site<br>40-43;<br>Ck2_Phospho_Site<br>12-15; Myristyl<br>22-27;<br>Pkc_Phospho_Site<br>36-38;37-39;           |                                       |
| DEX0305_179 |                                |                                 | Asn_Glycosylation<br>59-62;<br>Pkc_Phospho_Site<br>45-47;  |                                       |
| DEX0305_180 | 56-68,1.01,13                  | 1,o34-56i                       | Amidation 56-59;129-132;132-135; Myristyl 22-27;39-44;43-48;79-84;83-88;117-122;118-123;129-134;                           |                                       |
| DEX0305_181 |                                | 1,o15-37i                       | Camp_Phospho_Site<br>73-76;74-77;<br>Myristyl 49-54;51-56;52-57;62-67;63-68;66-71;<br>Pkc_Phospho_Site<br>2-4;70-72;71-73; |                                       |
| DEX0305_182 | 34-59,1.07,26                  | 1,i13-31o                       | Amidation 50-53;<br>Ck2_Phospho_Site<br>6-9; Myristyl 47-52;64-69;   |                                       |
| DEX0305_183 | 23-58,1.19,36                  |                                 | Camp_Phospho_Site<br>26-29; Myristyl<br>59-64;67-72;<br>Pkc_Phospho_Site<br>24-26;25-27;29-31;                             |                                       |
| DEX0305_185 | 6-15,1.21,10                   | 1,o20-42i                       | Glycosaminoglycan<br>12-15;<br>Leucine_Zipper<br>29-50;<br>Prokar_Lipoprotein 17-27;                                       |                                       |
| DEX0305_186 |                                | 1,i23-45o                       | Ck2_Phospho_Site<br>9-12; Myristyl<br>31-36;   | 42,.983,.69                           |

-128-

| DEX ID      | ANTIGENICITY                   | TRANSMEMBRANE                   | PTM   | SIGNAL PEPTIDE                        |
|-------------|--------------------------------|---------------------------------|---|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length | Predicted<br>Helix,<br>Topology | PTM   | Position,<br>Max Score,<br>Mean Score |
|             |                                |                                 | Pkc_Phospho_Site<br>16-18;  | ..                                    |
| DEX0305_187 |                                |                                 | Pkc_Phospho_Site<br>20-22;36-38;  |                                       |
| DEX0305_188 | 79-<br>97,1.09,19              |                                 | Asn_Glycosylation<br>42-45;<br>Ck2_Phospho_Site<br>23-26; Myristyl<br>41-46;<br>Pkc_Phospho_Site<br>69-71;93-95;                                |                                       |
| DEX0305_189 | 27-<br>36,1.02,10              | 1,o4-26i                        |   | 23,.99,.82<br>9                       |
| DEX0305_190 | 7-<br>39,1.02,33               |                                 | Camp_Phospho_Site<br>25-28; Myristyl<br>30-35;  |                                       |
| DEX0305_191 | 54-<br>72,1.2,19               |                                 | Pkc_Phospho_Site<br>5-7;25-27;  | 21,.881,.6<br>97                      |
| DEX0305_192 | 31-<br>47,1.01,17              |                                 | Asn_Glycosylation<br>33-36; Myristyl<br>59-64;62-67;<br>Tyr_Phospho_Site<br>26-34;  |                                       |
| DEX0305_193 |                                | 1,o22-41i                       | Camp_Phospho_Site<br>42-45;   |                                       |
| DEX0305_194 |                                |                                 | Amidation 43-46;<br>Myristyl 4-9;<br>Pkc_Phospho_Site<br>19-21;   |                                       |
| DEX0305_195 |                                |                                 | Asn_Glycosylation<br>25-28;35-38;36-<br>39;<br>Ck2_Phospho_Site<br>2-5;   |                                       |
| DEX0305_196 | 60-<br>72,1.17,13              | 1,i44-61o                       | Amidation 6-9;<br>Camp_Phospho_Site<br>81-84;<br>Ck2_Phospho_Site<br>107-110;<br>Pkc_Phospho_Site<br>84-86;85-87;<br>Tyr_Phospho_Site<br>68-74; |                                       |
| DEX0305_197 |                                |                                 | Pkc_Phospho_Site<br>40-42;  |                                       |
| DEX0305_198 | 51-<br>61,1.14,11              | 1,i20-42o                       | Camp_Phospho_Site<br>75-78; Myristyl<br>49-54;56-61;<br>Pkc_Phospho_Site<br>11-13;12-14;  |                                       |
| DEX0305_199 |                                |                                 | Camp_Phospho_Site<br>43-46;98-101;<br>Myristyl 19-<br>24;107-112;108-   |                                       |

-129-

| DEX ID      | ANTIGENICITY                     | TRANSMEMBRANE                   | PTM   | SIGNAL PEPTIDE                        |
|-------------|----------------------------------|---------------------------------|---|---------------------------------------|
|             | Position,<br>AI Ave,<br>Length   | Predicted<br>Helix,<br>Topology | PTM   | Position,<br>Max Score,<br>Mean Score |
|             |                                  |                                 | 113;109-114;110-115;120-125;121-126;122-127;<br>Pkc_Phospho_Site<br>58-60;95-97;101-103;  |                                       |
| DEX0305_200 | 10-30,1.02,21                    |                                 | Camp_Phospho_Site<br>22-25;<br>Pkc_Phospho_Site<br>21-23;25-27;   |                                       |
| DEX0305_201 | 5-31,1.02,27                     |                                 | Ck2_Phospho_Site<br>17-20;59-62;68-71;102-105;126-129; Myristyl 53-58;<br>Pkc_Phospho_Site<br>14-16;26-28;68-70;102-104;            |                                       |
| DEX0305_202 | 14-30,1.28,17<br>100-115,1.05,16 |                                 | Camp_Phospho_Site<br>36-39; Myristyl<br>115-120;<br>Pkc_Phospho_Site<br>18-20;35-37;87-89;110-112;                                  |                                       |
| DEX0305_203 |                                  |                                 | Ck2_Phospho_Site<br>18-21; Myristyl<br>36-41;<br>Pkc_Phospho_Site<br>33-35;   |                                       |
| DEX0305_204 | 13-26,1.04,14                    |                                 | Amidation 16-19;<br>Pkc_Phospho_Site<br>13-15;14-16;  |                                       |
| DEX0305_206 | 125-134,1.17,10<br>16-25,1.05,10 |                                 | Asn_Glycosylation<br>99-102;<br>Ck2_Phospho_Site<br>41-44;142-145;<br>Myristyl 13-18;34-39;<br>Pkc_Phospho_Site<br>149-151;206-208; |                                       |
| DEX0305_207 |                                  | 2,o10-29i42-64o                 | Ck2_Phospho_Site<br>32-35; Myristyl<br>10-15;81-86;   | 58,.927,.6<br>17                      |
| DEX0305_208 | 164-174,1.19,11                  | 1,o67-89i                       | Amidation 26-29;<br>Ck2_Phospho_Site<br>169-172; Myristyl<br>20-25;175-180;<br>Pkc_Phospho_Site<br>165-167;169-171;                 |                                       |
| DEX0305_209 |                                  | 2,i7-29o44-63i                  |   |                                       |

**Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

-130-

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. See, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1  
5 through 115. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). See also Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4  
10 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156  
15 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99  
20 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in  
25 combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and  
30 translocations. These alterations are used as a diagnostic marker for an associated disease.

**Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**



-131-

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

#### **Example 8: Formulating a Polypeptide**

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous

-132-

subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

5           Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation  
10 auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

          The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable  
15 polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl  
20 acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP  
25 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

30           For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

-133-

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

10           The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine  
15 or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium;  
20 and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

25           Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

30           Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture

-134-

is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

**10 Example 9: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 µg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

**Example 10: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

**Example 11: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and

-135-

separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is

-136-

removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector  
5 that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

**Example 12: Method of Treatment Using Gene Therapy-*In Vivo***

10 Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a  
15 promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7  
20 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers  
injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide  
25 constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention  
30 may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain

-137-

sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have

5 shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach,

10 intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by

15 the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or

20 less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05  $\mu\text{g/kg}$  body weight to about 50 mg/kg body

25 weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being

30 treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the

-138-

nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

#### Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection



-139-

(Paterson et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al., *Biotechnology (NY)* 11: 1263-1270 (1993); Wright et al., *Biotechnology (NY)* 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, *Mol Cell. Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993)); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," *Intl. Rev. Cytol.* 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., *Nature* 380: 64-66 (1996); Wilmut et al., *Nature* 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., *Science* 265: 103-106 (1994)). The regulatory sequences

-140-

required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

#### **Example 14: Knock-Out Animals**

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., *Nature* 317: 230-234 (1985); Thomas & Capecchi, *Cell* 51: 503-512 (1987); Thompson et al., *Cell* 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of

-141-

the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another  
5 embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive  
10 targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to  
15 express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial  
20 cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or  
25 transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered  
30 cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or

-142-

vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

5 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

10 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

15 All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments,  
20 which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

-143-

## CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
  - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes  
5 an amino acid sequence of SEQ ID NO: 116 through 210;
  - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID  
NO: 1 through 115;
  - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid  
molecule of (a) or (b); or
  - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic  
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
molecule is a cDNA.  
15
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid  
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a breast specific nucleic acid  
(BSNA) in a sample, comprising the steps of:
  - (a) contacting the sample with the nucleic acid molecule according to claim 1  
under conditions in which the nucleic acid molecule will selectively hybridize to a breast  
specific nucleic acid; and
  - 30 (b) detecting hybridization of the nucleic acid molecule to a BSNA in the  
sample, wherein the detection of the hybridization indicates the presence of a BSNA in  
the sample.
7. A vector comprising the nucleic acid molecule of claim 1.

-144-

8. A host cell comprising the vector according to claim 7.
9. A method for producing a polypeptide encoded by the nucleic acid molecule  
5 according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.
10. A polypeptide encoded by the nucleic acid molecule according to claim 1.  
10
11. An isolated polypeptide selected from the group consisting of:  
(a) a polypeptide comprising an amino acid sequence with at least 60%  
sequence identity to of SEQ ID NO: 116 through 210; or  
(b) a polypeptide comprising an amino acid sequence encoded by a nucleic  
15 acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115.
12. An antibody or fragment thereof that specifically binds to the polypeptide  
according to claim 11.
- 20 13. A method for determining the presence of a breast specific protein in a sample, comprising the steps of:  
(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the breast specific protein; and  
(b) detecting binding of the antibody to a breast specific protein in the sample,  
25 wherein the detection of binding indicates the presence of a breast specific protein in the sample.
14. A method for diagnosing and monitoring the presence and metastases of breast cancer in a patient, comprising the steps of:  
30 (a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient; and  
(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the breast specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the

-145-

polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of breast cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said  
5 kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient.

16. A method of treating a patient with breast cancer, comprising the step of  
administering a composition according to claim 12 to a patient in need thereof, wherein  
10 said administration induces an immune response against the breast cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the  
polypeptide of claim 11.

15

11

12



## SEQUENCE LISTING

<110> diaDexus, Inc.  
 Salceda, Susana  
 Macina, Roberto  
 Hu, Ping  
 Recipon, Herve  
 Karra, Kalpana  
 Cafferkey, Robert  
 Sun, Yongming  
 Liu, Chenghua

<120> Compositions and Methods Relating to Breast Specific Genes and Proteins

<130> DEX-0305

<150> 60/268,999  
 <151> 2001-02-15

<160> 210

<170> PatentIn version 3.1

<210> 1  
 <211> 357  
 <212> DNA  
 <213> Homo sapien

<400> 1  
 cgggccggca gtatgatgga tcggccgccc gggcaggtac agctgggtccc actcctctct 60  
 ggtgaagtc acggccacga tcctgaaacg tcagtgttc ctgagatctc accatctgtg 120  
 agccatcatt cattttcttc tcctccatgt tcccctctcg agaaaaaaca gcattctgag 180  
 aaggcataac ttcttttttg agtctctcga ttcagtcttc cactgggatt acacctctct 240  
 gcagttctta tggtgtaatg tcgccaaagc tctgctatct tctacatgaa agtcagcaga 300  
 tgcaccagga ccagcagctt aaggagctgg ggctgctctt gaaagttgat gtccagt 357

<210> 2  
 <211> 2152  
 <212> DNA  
 <213> Homo sapien

<400> 2  
 agcggagcgt cttgcgccgc cattgcgggg aggctgtcct cagagcaggt ctggcgcgcc 60  
 ggtggctgga ccggccccag gagcccagtc accgggcgtc attgggtcag gctgcggggc 120  
 cctcggcacc ttctccctcc cgggtccacc gcggcgcgcg cggcgggcggc ggcggcgacg 180  
 gcggcgcgct caggtggcgg agcctgccga agcgcccttt gtctgcggag gtcaacatac 240  
 ctggcctaag gaggcaggat tgagtgactc tcaactacca ctggtgttgc tctttgaaag 300  
 tggcgcttgg caccagcatg aactcccat cctcagcaat cccatcaggt gttttgggtc 360  
 ttcaacctaa aattctatct tacaagatcc ttgccaggat gcagatttga atactatagt 420

2

```

gaagtctgta catgaagaaa tgatgctttt agggaggaaa aaacaacaag gtaataacaa 480
ccttcaagag ccccttcata tcaactcggc ataaacaagg caagattctg agagtggccg 540
cccctggaag cagaaattat tctgtgtggc tatccatgtg gctcctgagg ctctaatacag 600
agatggggca cctttagtac caggggagtg actgttgccc ataaggtact ggacatcaac 660
tttcaagagc agccccagct ccttaagctg ctggctcttg tgcactctgt gactttcatg 720
tagaaagata gcagaagctt gtggcgacat tacaacataa gaactgcaga gaggtgtaat 780
cccagtgga gactgaatcg agagactcaa aaagggaagt atgccttctc agaatgctgt 840
tttttctcag gaggggaaca tggaggagga agaaatgaat gatggctcac agatggtgag 900
atctcaggaa tactgacgt ttcaggatgt ggccgtggac ttcaccagag aggagtggga 960
ccagctgtac cctgccccaa agaacctcta tcgagacgtg atgctggaga actacaggaa 1020
tctagtgtca ctggggatc agctttgtaa gccagaggta atcgcgagtg tggagctaga 1080
ggaagaatgg gtgatagaaa gagacagcct gctggatact catccagatg gagaaaacag 1140
acccgaaatc aaaaagtcaa ccacaagcca gaatatctct gatgaaaatc aaacctcatga 1200
gatgataatg gagagactcg caggagacag cttctggtac tccatcctag gaggactctg 1260
ggattttgat taccatccag agtttaacca agaaaaccac aagagatatt taggacaagt 1320
aactttgacc cacaaaaaga tcacacagga gagaagcctt gagtgttaata aatttgaga 1380
aaactgtaat ctgaactcaa accttatgca gcagagaatt ccttccatta aaatacccct 1440
gaattctgac acacaggga acagcatcaa acataattca gacttgattt actatcaggg 1500
aaattatgta agagagactc cctatgaata tagtgagtgt ggaaaaatct tcaatcaaca 1560
tattcttctt actgatcata ttcatactgc agagaaaccc agtgagtgtg ggaaggcctt 1620
cagccacacc tcactcttta gccagcctca gatgttgctt acaggagaga agccctataa 1680
gtgtgatgaa tgtggaaaaa gattcagcca gaggatacat ctcatcaac atcagagaat 1740
tcacacagga gaaaagcctt ttatatgcaa tggatgtggg aaagccttcc gtcagcatc 1800
atcctttact caacatctga ggattcatac tggagaaaag ccctataaat gtaatcaatg 1860
tggtaaagct tttagccgca tcacatccct tactgaacat catagacttc ataccggaga 1920
gaaaccttac gaatgtggtt tctgtggcaa agccttcagt cagaggacac atctgaatca 1980
acatgaaaga actcatacag gagagaaacc ctataaatgt aatgaatgag ggaaagcctt 2040
tagccagagt gcacacctta atcaacacag gaaaatccat actcgggaga aattatgtga 2100
atataaatgt gagcaaactg ttcgccacag tccttcattt agcagcacat aa 2152

```

```

<210> 3
<211> 1079
<212> DNA
<213> Homo sapien

```

<400> 3  
 acaaattata cataataaag tgtttttaat aatcaaaaaa aaaaaaaaaa aaaaaaaaag 60  
 acctaaaaaa aggggggttc aaaaaattgt ggcaaaacac tttctcaagt caataggcca 120  
 accccattca cccattactc gggaacaaag gtcccgaagc acgaagggca aactcagcga 180  
 tgcattgcagg caacaggaca aacaaaggcg gtgaaaagcg aaaagcagaa agacgtacag 240  
 catgcagtga tcgaacaacg gccaaagaaac gcgcacaatg gtggcggttc agcatggcag 300  
 gcagcgcattg ccacacgcgc ggtcacaagc ggaatccaac gacagcgcac agaaggacgc 360  
 cgaaggagca gacatatcca cccagagca aataaatcaa cgcttgccga cccacaggag 420  
 caaaaaaacc tacaacgccg caaacgacac ccactgctcg ctccatgggt gggcaccacg 480  
 agaaacaaca cccctagtgt acggaaacct ctgcacccc gccaacgaca gggcagactc 540  
 tggggcacga caaactgcc aagcaaaaga aagcgcccca catcaaatg aggaccaagt 600  
 cggcgaaaaa acaccccgat agtggggcac acaggcacca acagaaccag ctgcatgccg 660  
 tggcgcacca agacgggtcc gccgggtggg cgaacaatca cggggggcaa gttggggacc 720  
 aacaaaacac acctgcggat ggggggccct cccctgttag gaccacgtat tatgatactg 780  
 aactacgagt atcaacaag tagtaaagac ctaacatact gaaaatcact atatagtgt 840  
 ccgagccgag tgggtgtgtc cacataccta gtgcgacaca cggctgtcga tactacgaag 900  
 tttgatagat caccggaacg agcttaccta tacatatggt aaacccgtac ggtggtgtga 960  
 gtgattggaa ctcatggagt gagtgatctg ccccgtaaaa ataattctag cggagaaaca 1020  
 gttgtccgag cgaaacgcag acatctgttc acagctgtgt ctggacaaac aacttgta 1079

<210> 4  
 <211> 348  
 <212> DNA  
 <213> Homo sapien

<400> 4  
 atggagagct cccgcgtgag gctgctgccc ctctgggagc ccgccctgct gctgatgcta 60  
 cctctgttgg gtacccgtgc ccaggaggac gccgagctcc agccccgagc cctggacatc 120  
 tactctgccg tggatgatgc ctcccacgag aaggagctga tcgaagcgct gcaagaagtc 180  
 ttgaagaagc tcaagagtaa acgtgttccc atctatgaga agaagtatgg ccaagtcccc 240  
 atgtgtgacg ccggtgagca gtgtgcagtg aggaaggagg caaggatcgg gaagctgtgt 300  
 gactgtcccc gaggaacctc ctgcaattcc ttcctcctga agtgctta 348

<210> 5  
 <211> 1782  
 <212> DNA  
 <213> Homo sapien

4

<220>  
<221> misc\_feature  
<222> (322)..(322)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (466)..(466)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (469)..(470)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (474)..(474)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (1287)..(1287)  
<223> a, c, g or t

<400> 5  
ccccccctt cttttttttt ttttttgga tttttatga ctctttattg gaaacagggt 60  
ctcaatcttg gtcactctca ggccagagggt gcagtgggggt ggtgaccaca gtctccctgg 120  
tcagccttca acctctccca agcttcaaac aattcctctc ccaccttcag ttcttccaga 180  
agtttagcgtg ggactacgag gtgtgcaaca acaccattac ccgagggtga atttttttgt 240  
gtggcagaaa taagtgggtcc tcagtgtgtg ttatctcccg agggcgtgggt gtgttaaaaa 300  
actctctcgt ggagcctcga gnggcgaact cctctcccgc gtgtctctgc ggccctctcc 360  
ccaaagagtg tggctgtggg cggattacga ggggtgtgtg gacgcacacc ccgatgtggc 420  
gcgtggggct atatatgtgt gttttctttc tacaaaaaat ctatangann aaanatctct 480  
ctcccgagag atgtgtgtct tacacaaaaa cctatatgcg ggcaccatat atttctctta 540  
tatatattta taccacatg tgcacatttg tgcacaagga aaatatcttt ttataaaggc 600  
tgtgggacga gggagatata atattaagag agagagaggg gcatttattt ctaaaaaacc 660  
atttaaggag aggcgcgtgg gcgaaggcta taaaagaggc gaaggaaaac tctccagggg 720  
cgggcgacaa acattattta tctgcggtgt cctataaaaa aatttcttta tgtgtctttt 780  
ttacgaaaaa gagagagaaa caacaccaag aggcgcctgt gagagggcga tctccgacgg 840  
gtgagacctc cataaaaaag ctcttctccc caatcttctt tcaagaggaa aaaagggcgt 900  
ggaacaatat agcgcgttat aaatctcttt ataccccaaa gaggaaaaaa cttcgaggaa 960

5

```

aagaggcgaa tttttctcta taagtgggtgt ttctccccc aaatgcggcg cttacacacg 1020
ggagccagac gggaggccaa aaaactcccc caatattttc aaatctgtca gtgggacaac 1080
cacaacagcc ttccaatatt taaacctctt gtgggcgttg tggaaggggg cgtctttttc 1140
aacacaaaaa tttgtgttgt tttttaaaact ttttttttcc caaggggaac accttaaaac 1200
gacgttgggc cggaataaac ccaactgggc gcaaataagg ccgtgggtcc tcgttggttg 1260
ctcggaaaac ctctcggctt aatcccnggt ttcacaaatc cccacataag aagatgagaa 1320
ccacgcacac gcagaactaa ccatggcacg ggacgacacc gcacgaacga caggacacga 1380
cacgagcgaa cgacgcgcc acaccacaca cgcgcaaacc agacagacga gacagcagag 1440
gaccacacaa cagatagtgc caccaatcaa aaacgaaacc agcaaccaag actatataaa 1500
gcagacgaac acatacagac ggaccaccaa gacacgaagc agaagaacga cgaccaagga 1560
gaacaaagac gacaaccgaa gacacgactg caaacacag agtgcagaca caacgaacac 1620
acgtaaacac cagaaccgac acccaacaca acaacagaag ccacagacag caaccacaca 1680
gacaagaaca gcaaacagac cgaacacaag cacacaaaca gacaaaatcg acatcaacac 1740
acacacagaa gacaaaacaa cagcagaaac acagccaagg gg 1782

```

```

<210> 6
<211> 1023
<212> DNA
<213> Homo sapien

```

```

<400> 6
cggctggccg gggagggtccc ccccccttt ttttttttg gtttttttt ttttttttt 60
tttttttttt ttttattatt atttgaaaa aaaaaaatt ttttttatat ttcatatattg 120
ttggtagggg ggggtgaaaaa gaaagaaaga aaaagagaga gaaaaagatg agaggggggt 180
ggtgatggag agaacaataa ataaacaaca taatggagta gagagtgaaa cgtggtgtgt 240
gttggtgtat catacacacg catatctca cggagggtgc acactaagaa ccgacgtaca 300
ttgtagatga gatagagaca tcaacacatg aagaagtgtt gatgatacgc gatagacaac 360
acaacaatga tgaagcacac acacactaca catctaccag aacacagaac caccacgaag 420
acaaaaacac gcgacacaga cacaacacac agagcgaagg aggaggagcc acccacaaaa 480
actcgccaca cagcagcgcg tcacatcaca acaccaccac cagcagcgca ccacagaaga 540
gaaaaatatg aatccagaac aggcacaaca taagaagga taagagacac agaacaacgc 600
atgagagaga aacacaacag gaggcgacaa caacctgacg aagacacgca acgcgagagc 660
aagaagccac agcaaagtag cgcgaaaacg actcaaacac acaaagtcac cccctaccac 720
cacgaccact ctccgacaac acagcacagg aaagacaaga acgtcaaggc tcgagaccaa 780
ccacacccaa acatcgctga aaacgacgag acaccacaaa aagtaaataa catgatgaaa 840

```

## 6

gacaaacaca acaaagcaaa gcctaacacg aaacaagcaa aaaaaggaaa gaaaaacaga 900  
 cacgactcag acagcagaag taccaaaaga ataagacgca agcagatcaa gacaaccgac 960  
 agatagcgaa gtcacacgga aaagaaaaag taggagagaa gagacagcca aaagatacaa 1020  
 gga 1023

<210> 7  
 <211> 35  
 <212> DNA  
 <213> Homo sapien

<400> 7  
 acccaatttt atatcctttt ttaaaggagt gacct 35

<210> 8  
 <211> 540  
 <212> DNA  
 <213> Homo sapien

<400> 8  
 cggcgccggg caggtaccgg ggactacagg tgcattgccac tacatccaac tagttaattt 60  
 tttttttttt tttttttttt tttttttttt ttggaaaaag ggggtcaaat tttggtggcc 120  
 cggggtggtt aaaacccgtg tgggtcaaga aattttccca cgcttggccc tccaaagagg 180  
 tgctatagga atatacgggg tgggaagcta taccactttt tgtaggaaat ataacaattt 240  
 attttattta attaatataa aaaaaagtgt ctccatgtgg gcaacagtgg tgttactcag 300  
 gcagaagaaa aagcgactt agaagtgtga gggacctata aaacaaattc gagtgttgac 360  
 agggatttct ataggagct atacgctttc tgctaataata ttatttactg ttgaaaccag 420  
 aaggattggg ggcggtaaac taagtgggc aacagtaggc tggtttgtcc gtggttgag 480  
 aaagtagtgt atactcgcg ttctaaattt tccacaaaaa tattagtga agaaaggaga 540

<210> 9  
 <211> 645  
 <212> DNA  
 <213> Homo sapien

<400> 9  
 gcggcgccgg gcaggtacag ttgttcctca ccattgacctt ggggtcccg cccaactaat 60  
 ccagagccat gtgggtttgc agagacaggc attcctccca tattctggcc tctgacctga 120  
 aatcttctaa cttgagaaga gaacagtcac cttcctggga atctgaaata gaaaggcaaa 180  
 tttgtgaagg cttttctgac atctgaatgg ctggatttgc atttctgta gtgataactc 240  
 agtgccatcc agacctgaca gtgatgaacg atgctggatt ctgctcaaat tccatcaagg 300  
 cctgcagggt gaagactctg gtccctgaac ccagtgtcct ctggcccttc ctgtcaaagc 360  
 attggagtga caggagagaca tttgagaggc agtgaggagg aaggacagag gcatcagggt 420

7

```

gggtgtggca gcttccatat ttacgcacgg gcagaagcag cagatgaggg taagattcat 480
gagtgaggaga ggagggacgg ttagagaaca atgggaaaat ttccttcttc atgtaagaat 540
ctggacctta ttgaagtctc tctgcttgt tgggcaaaag taatgaaact ccattggctt 600
cagatgaggt cactccaatg atcacagcat aaaaagatca ctcaa 645

```

```

<210> 10
<211> 806
<212> DNA
<213> Homo sapien

```

```

<400> 10
gcggcgccgg gcaggtacag ttgttcctca ccatgacctt ggggtcccg cccaactaat 60
ccagagccat gtgggtttgc agagacaggc attcctccca tattctggcc tctgacctga 120
aatcttctaa cttgagaaga gaacagtcac cttcctggga atctgaaata gaaaggcaaa 180
tttgtgaagg ccttctctgac atctgaatgg ctggatttgc atttgctgta gtgataactc 240
agtgccatcc agacctgaca gtgatgaacg atgctggatt ctgctcaaat tccatcaaag 300
cctgcagtgt gaagactctg gtccctgaac ccagtgtcct ctggcccttc ctgtcaaagc 360
attggagtga caggagagaca tttagagagg agtgaggagg aaggacagag gcatcagggt 420
gggtgtggca gcttccatat ttagcagaga gaagcagaga tgagggtaa attatgagt 480
ggagaggagg gaaggttaga gaacaatgga aaaattttct tcttcagtgt aagaattctg 540
gacccttatt tgaagtctct cctgcttgt tgggcaaaag taatgagaaa ctccacttgg 600
cttcagaatg cagtgtcaac tccacatgaa tcaaagcaat aaaaaagaat caactcagag 660
caggctgagc tatgtgaggt atgaaaactt gatcagggcc agcgtgagta tgggacttca 720
gtcatgtccc cactccctca caggaccac acgggtggag ggtgggggga attgtttaaa 780
agcatttagt tcctaaacta gctgcc 806

```

```

<210> 11
<211> 122
<212> DNA
<213> Homo sapien

```

```

<400> 11
ccgaggttgg gtatccttgt tactgattgc catggaaatg cctctagatg tgtctccatt 60
aagagagcgg ctttagaact taacacaggc tgccggtgct ggtgaaatac ccatcaacgc 120
cc 122

```

```

<210> 12
<211> 861
<212> DNA
<213> Homo sapien

```

```

<400> 12

```

8

```

cgccccgggca ggtaccagac gtggcaaatc tcaagtgaca gtggaccccc ccccccgcg 60
ccagttaata aattcgcccc tttttccaaa ctttcccagc atcagcatcc agaggtcagc 120
aggaagcttg agttcattat accttccttg gggtgaccct cccacacca atctctttgc 180
tctcacttgg gaacccggtg tgcctcacgt ttatatctta actatattgc aattatgtta 240
cattacattg gttttggtat tccaagctag cctctggggg ttaaatctag tcgccacggg 300
gcccttggtc ctttctcttg tatacactat ctaccaggtt tgtggattct atcatttata 360
caaatattat tgcttgctgc cgattctgtg gatttcttat actattcgtg tcgggcgtgt 420
gcgctgtgaa attaaacttg cgcagacgac tctcacaact acttctgcag ggcgtgacta 480
aggtggtcga caaacacaaa attagccaac gatattgtga gacctcaca gggtttacca 540
cttctctcaa acccgatgag tgttacattc acctgtggcc acctttataa gcaatgtagc 600
ttcaactcaa acggggctct tacatacggg ggggggaaaa agacaacacg ctccaactgg 660
tcttggtgga acaataactc acctctgctg ttgaaccatc cttatgcagc gggccatgtg 720
ttgcgggctc cgtgaaaacc aacgcttcgg ggaaacacct tgggggtttg gacgcagaac 780
ttgcgggcca tccccgcgga caaacggcc tgaaattgta ggaaatccc gggaaaggcc 840
ctgggatccc cgcattaaac c 861

```

```

<210> 13
<211> 1009
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (782)..(782)
<223> a, c, g or t

```

```

<400> 13
ccccccccc tttttttttt tttttttttt tttttttaag agaaaaaccc ggaaatgatt 60
tcgggggtga ggaataggag aaaaatgggg aaatagggtg gttattaaac attgaggggt 120
gttttcctcg gtgtggaatg agggtaaagt ggtggtcaag tgggtgggtg gctgtagttg 180
acccccatgt gtgtgtgtg ggtggataaa atttgttaaa gggatatata gggcgtggaa 240
catagtatat gtgtgtggag ctccgtgtta agttagcgaa aagtgtgata tattgtggat 300
ctcacggaaa aagtgtgtgg gttccatagc cacaaggaga agtttctctc ccaggatagg 360
ggttaaaata gggggggggg ataaggcgga gatttatagc gcaagaggtt gtgtccataa 420
aaaagtttct tgtccaaaga aggcttatta tgagagcggg gacagatcta aaaagctttt 480
gtgaaaagat ttcccttttt aaggaaaaag agggaaattt ttgatgaatg tggcaaccag 540
ctgtgtgtag aagagtggcg cgttcgcggg aaagcagtgg ggagattttg ggtccttaag 600

```



9

```

gggacgacac acatatcagc ttccacagcg cagagaaat gtgttttaaa agccacgccg      660
gggagggggag acgcgacaca aaataagctt gaagcaaaaa tatgaaaata agtggtgggc      720
tcgccgagat ttagaacaag cgcggggggg gagggagaaa aaaaactccc gatgtgtggg      780
cngccccaca taacggaccg tgggtgtcac ccgcgggggc ggggtggtgc gcaccaccag      840
ttggcggggtt atacatcccg cgggcgcccc caaaaaattt ttccccacac aatataatTTA      900
gtcgtagcag ccacgtacaa acaaccaaac ttaggtgtac acgagacgag acacacacac      960
aaaccaccca ccaccagcaa caacaacaat caagacacag acagaaaga                    1009

```

```

<210> 14
<211> 357
<212> DNA
<213> Homo sapien

```

```

<400> 14
taaaaaatta tttgtagaga tgggtctcc ctttgtgctc aggtctgtcc tgaattcctg      60
gcctcaagca gtcctcctcc ctccagcctcc caaagtgtcg ggattacaga tggtaagcca      120
ccacacctgg cctttttaaa caacttctga gactagggtt cctcatagtg gcatatagaa      180
tctttcatag atggctgcag caatgtctcc cattccactg gccttcagtg accttgccac      240
ttcttcatca agaggtagag tctcttacca ccctgccttg catctgggca gtccctgtga      300
ttactttgat cagtagcata cagtggaagt gatgggtgcc actactagac aacactg      357

```

```

<210> 15
<211> 415
<212> DNA
<213> Homo sapien

```

```

<400> 15
ggttggttat ttacaatgca tgggccagcg tctccttggt cttttccgct gtccccgggc      60
gcgtacagtg tgcaccagca gtactgagtc acagttccaa ccagatctcc taaagtgtgt      120
gaccaaaggt gttgctgagt ttgaacacat tgcataatTTA aagttgcaa tagcgacgat      180
gtgggtgagc aagttagatt acttttgtct ctatgggaca gctttgacct attctccttc      240
ttggtcttcc cagctggggc attcgtgcct atagtgttca agcagtgttc tagaggaaat      300
taataagttc tgaattccta ctgtacacta atctcctctt ccacacacct ccgggtctct      360
cttaacttga ccctcatggg gctacactac caciaaggca acatctctcc ttagg      415

```

```

<210> 16
<211> 893
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (516)..(516)

```

<223> a, c, g or t

<220>

<221> misc\_feature

<222> (788)..(788)

<223> a, c, g or t

<400> 16

```

tttttttttt tttttttttt aaataaaaaa aagagtccac ctttaaatta ttccgtggtt      60
aaacactttc ttatgggttt gggtaaaaag caatattcta tgaccttgaa atagtagtag      120
gtggataact atggtggact ggccagtaaa ttctctatta ttttcaacgt ggtgccactt      180
agagagataa tcgaaaatat tatggaacta ttatttatat ccctaaaatg tccctggcaa      240
taacacctac tatcaccacc ttaaaactatc ccacgctttt ataaacacag gctttttggg      300
caatatcacg ctatggtgga acgaaatgtc tcacgcgcgt ggtgcgcctt acatattttg      360
gatatggggt atttcttcg tggegccttc atattctggt atctctctca cacacacag      420
caaaacacag tgtattcatg ggggtggcctc tcttccttac aaacacagca cacacctggt      480
ccccccctat ttgggggcgc atatatatta tctctcttc catagcgtgc gagtgcgcgc      540
gagtataacg cgaaagcctc ctccagagac agaacaaccc ccacttcgag gcccgcgggg      600
gcgccaagtg tgggggtgtg ccaaactccc gtgcgactcc gggtagcgtg caccctatct      660
ccgggcgctc cctgtactta tatattccct tgtaaaaagc gcacctcagg ggctccccct      720
tttcatacaa cgacgcccac gctccctcca caccgcgctc tggggcagct gcgagacgtg      780
ccccctntc ccggttggt tgcttcgacc tccccccaa tttttatct cccccactt      840
tgggttgcca ggctccccac ttcacctcct cgggcgccta ccctacattg gcc      893

```

<210> 17

<211> 458

<212> DNA

<213> Homo sapien

<400> 17

```

gcgtggcgcg gcgaggatga gcaagactcc tgtctcacag aacatgacat aagataaaat      60
acaataagta acagatgtta ttttttaaaa agctaacttt atttaataac tataacgaca      120
cagaaagatg cctttctcac actgaatctt caagattcta aggaagaaca tacgagtctc      180
ctttgcgaat gtccaagtaa gtaagattta gcacggaaat ctaatcaagc atctacttgt      240
cctcacatgg aaatacttat gaaacttctt ataagagagc agtaatctct aggccggaca      300
ctgttgcat catacctgta atcccagcaa ttttgggcag gccgagggtc ggtggtatca      360
cttgtgacgc tcagagaatt actaagcacc agcctggcgt atatatggca atgagcctcg      420
aactctatct agaagaatta caagaaacga aaagaata      458

```

<210> 18  
 <211> 542  
 <212> DNA  
 <213> Homo sapien

<400> 18  
 ccgggcaggt cccctcccct tttttttttt tttttttttt ttttggaatg aaactggcaa 60  
 tttttataaa aaagttataa aaattaaaaa aaaaaaccaa gcctataccc aaaacacaaa 120  
 aagcaacgac acacaacctt tctcagagtt ttactttacc tttgtggagc gttcacacac 180  
 ttatttacc acccttagtg gcttttttta aaaatgggtc aatttctcag ggtataggag 240  
 ggagctgtga gtctcgggta taatatgagc gccagcccat ctacagaggt gttacctata 300  
 atttatagag tgctataaaa tataaacaca gggctccctc atttgtgaaa aaaagaagaa 360  
 aaggaaacac tattttccgg ggtgggggtt aaattttagg ccaatgggtt ataaaaaac 420  
 cttgggggtt aatctcaggc ctcatagcgt tgtttcccg ggtgtgtga aaatgtggg 480  
 atatctccc gcgctccacc aattctcaca ccacaacctt tccccggaa acaacaacg 540  
 ag 542

<210> 19  
 <211> 326  
 <212> DNA  
 <213> Homo sapien

<400> 19  
 tgacataata taagatagag tatagataga aagagaccgc ggggtgaaaca ttcaggaaga 60  
 tcacagagga aagaatctgg gaatagcaac agcacggaag ttgtttatag aatccgctag 120  
 gttatgccgc cctaactcta tatcatgcag attatgatcc tagtcacaat attgttgact 180  
 ttgaaaaccg aactatcaga tactccgttc aggcaccaga ctggctatga agtggcacat 240  
 acatggaata gacccaaata ggactgcgaa gatgttgaaa aataaactga cattagaaca 300  
 acatcccaa gagaggttg gacttg 326

<210> 20  
 <211> 603  
 <212> DNA  
 <213> Homo sapien

<400> 20  
 cgtggtcgcg gcgaggtact tagagtttct gtttgattct tttttaataa actactcttt 60  
 gatttaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaggggg ggggttccca aacccgggtg 120  
 atgtttggaa acagtccttc ggattggagg ggtttcacc ctgccaagggt gggaccaccc 180  
 aagcctcgtg tgacaacgcc ctcttaacag tgggaatgcg atcgacgcac gggctctgag 240  
 gatacttgcg cacagagcac actgactgcg atcgaatctg ggacttcagg gggctatcgg 300

12

```

tcgtggggag cctcgctctc ccttgggggc gccgcggggc tggteccact tactccccagc 360
gattcagaga aggcgacctt tctgggattt ctcacgcca cggagggatt ctccgtgagc 420
ttcgactgtg cactcattcg acacatttaa cagaacgaaa actctttttc tggccccaag 480
tctttttgac agggactgga aacagctggg gcagtaacct ccttggctca taegcctgta 540
ctcctggtgt cgaacttggg aaagtccggg tcacatattc cacaaaattt acgcaaacca 600
agt 603

```

```

<210> 21
<211> 513
<212> DNA
<213> Homo sapien

```

```

<400> 21
atggctaaat tcgtgatccg cccagccact gccgccgact gcagtacat actgcggctg 60
atcaaggagc tggctaaata tgaatacatg gaagaacaag taatcttaac tgaaaaagat 120
ctgctagaag atggtttttg agagcacccc ttttaccact gcctggttgc agaagtgccg 180
aaagagcact ggactccgga aggacacagc attgttggtt ttgccatgta ctattttacc 240
tatgacccgt ggattggcaa gttattgtat cttgaggact tcttcgtgat gagtgattat 300
agaggctttg gcataggatc agaaattctg aagaatctaa gccaggttgc aatgagggtg 360
cgctgcagca gcatgcactt cttggtagca gaatggaatg aaccatccat caacttctat 420
aaaagaagag gtgcttctga tctgtccagt gaagagggtt ggagactggt caagatcgac 480
aaggagtact tgctaaaaat ggcaacagag gag 513

```

```

<210> 22
<211> 136
<212> DNA
<213> Homo sapien

```

```

<400> 22
aagatagtgc cactgcactc cagcctggca acagagcgag acaacatcaa aaaaagtagg 60
aaggaagggg ggggaaggagg gagggaggga aggaatggaa ctatgactct aagatgctac 120
actctgagag tgtaaa 136

```

```

<210> 23
<211> 933
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (661)..(661)
<223> a, c, g or t

```

```

<400> 23

```

13

ccgggcaagg tctttttttt tttttttttt ttttttttgg agggaaaaac ccggtaatga 60  
 tttcgggttt agaggaatag gaggaaaatg gggaataggt tgtatgagaa catgagaggt 120  
 gtgtgtttcc tccgtggtag aatgaggagg gtgtttaatg tgttgtgtaa atggtgggtg 180  
 ggtgtgagat tggtagacgc cccattggtg gttggtgggt aaattattgt acgaggggat 240  
 gatataaggg gctggtggac tatgtattgt gagatgtctc tggaaatgtc agagagaagt 300  
 tatatatatt gtggtatcag agagagaaca gcgtgggtgt tctaagacc cacgagaaga 360  
 tatgtttctc ccacagagta gagtgttaaa taatgtgggg gggggtgtaa gaggcggaag 420  
 tgttaaagcg aagtgtcttt tgtcttaaga agatgtacta tcaaaacaag actcttattt 480  
 cgagtggggg aatgtaaaag tttggggaaa cgtctccttt ttgaagaaga gaggcggatt 540  
 tatgttgatg tgcgcaaact gtgtgtagag tgttgcggtg tcacaagaaa gtattatagg 600  
 aaagtgtgtg ggctattagg gcgagaaaca aatagtttac ctcgagaccg agaatgtaga 660  
 ntaacgcccc cggggggggc ggcccagcat gtataatcta gaaagaaata gtagatgttg 720  
 tggcgcgccg cagcgtgtag agacgacgtt gggcggggga tagcccaaca acgtcggcac 780  
 acaataagcc ggtgagtacg gccggggggg cgtgacagac gtcgggtgtt catctcacgg 840  
 ggcttcaaca aattcaccta ctacaactcc atccccacaa caaccacaca cacagctcaa 900  
 caacaccaac gagacgaaac aacaaacgaa cga 933

<210> 24  
 <211> 911  
 <212> DNA  
 <213> Homo sapien

<400> 24  
 ggcgcccggg caggtaccct ggtccagagg gtttgttctt attggagggc tatctgcacc 60  
 tctctttgaa tctcttgaa tagggagata aggagaagaa ggaaacataa attgatggct 120  
 atgcctgcc ttctccgttc tgcttatccc tggtaagggt tgccagagaa ttcaggccct 180  
 tcagagccag ctgagatgtg ctgatatgct aagtgattcc tcatctgatt ccttgctcca 240  
 gaactacagg gacttgaaga cagactacat ttttcctgag cgagacaatt tggcttcaag 300  
 ggaaacccaa actgtagcac agaattgtgag gtgagtttgc ccttgccctt tcatttatct 360  
 tcctttaatc aaacagacta aacgttttca ttggaacaga gaagattgtt atccttggtc 420  
 ttcttgtgtc tccagcagta tttttcttag gaatgtgtta atagctgtaa aaattttaac 480  
 acgtcttcaa gtgcctctca tgtaggaga ttcttctcag ttgcgggaaa agttgtgtgc 540  
 agattgcccc gtatttaacg tgaaatccca aatgtttctg acaggttgat tatgctcttt 600  
 cttcaaatgc cctgtctttt cagagtatgc agccagatgc ttccggaggg agagacattt 660  
 tttctttgcc aatcccgatt ccttcagtcc tcaatcactc ccagaaagt taggtccaaa 720

## 14

agacgggttaa ctttcagcga caagtaacga acacgattgg ggtggtctca cgggtcaagga 780  
 tagtgtgggtg ctggcctttc gtaacgagtt atttgctcgg tcaccaactc ctttacctta 840  
 atgtttggtc gaggaccaga acctttacgt acaatatggg tgtgtccgct taacggttca 900  
 aaaagttgca c 911

<210> 25  
 <211> 475  
 <212> DNA  
 <213> Homo sapien

<400> 25  
 ggaaaacaac tttttatgta tagcttctaa aaggaagaaa aaaaaaaaaa aaaaaaccct 60  
 tggacttcca cgtgcccac tcaagaacat tccactcaca gaattggagg ttctgggac 120  
 ccagggtctg ggagtttccc aattgggttaa ttggtaaaca ggaacggggc acacacacat 180  
 ttaagatgaa tggttaattat tatccctcct ggctgggtca ctaccggtcg cttctctatt 240  
 tctcttctct tgggtgtaatt ttatttaaaa gaaaaaaaaa cttttggtaa cgactattcg 300  
 gcaggtttaa aaatcaaata aaccccggtt tttttcaacg aaaaaaaca aaaaaaaca 360  
 aaaaaaaca aaagcgcgcg ggggggaacc cggggcgcaa aaagcgcggg tccccggggg 420  
 gagaaattgg gttccccggc caaaattcc ccacaaaaa agcggagaac aaagt 475

<210> 26  
 <211> 709  
 <212> DNA  
 <213> Homo sapien

<400> 26  
 aaaaaaggg taaattgggt aaaaattcag gtgggttagc aaacaaaaa ttaattgatt 60  
 aggaattggc aagtttggg atgtttccag gggatttctc agcctttaa ttattagaaa 120  
 cagcagaaat tttgtgaaa agtaaattat ttggaaaaa tgaattggca tgcagctagc 180  
 ctttgtgta ttaacaaata atttttctag atttgggacc cctaattagt ttaaaaattt 240  
 aaaaatttaa accattaaac attaggggcc ttttaaattg tgctcgggta taatattatt 300  
 aagaatagaa ggcttgaaac tgtggtggtt aagggtctct tcgtggtggg aagggtccca 360  
 tttacattct ttattattta cgtcaagggt ccattgaaaa ctaactgtgt ttaggatcgg 420  
 tctggaaatt ggctaagtct caggcagggg taaatcctgc tctcaggggc caacaggggg 480  
 ggaggcaaaa tagaaaacat tccccagata ataagctttt atcaattttt ggaggcaacg 540  
 atgggaggta actcagcgaa atattacgtg ggtcctgtaa aaggaattaa gggggaacgg 600  
 gaacattttt aatgggagga gaaattttct ttttaaaaag gccctaaaga aaatggttgg 660  
 tagaaattcg aattaatttt aacataattt ttgggttatt tcttaagga 709

15

<210> 27  
 <211> 722  
 <212> DNA  
 <213> Homo sapien  
  
 <220>  
 <221> misc\_feature  
 <222> (143)..(143)  
 <223> a, c, g or t

<400> 27  
 gccgcccggg caggtactac tgtaatatataa aaagtcactg tatttgcgat aaattctttt 60  
  
 ctattataaaa aaaaaaaaaa aaaaaaaaaa aaaacacaaa aaaaaaaaaa aaaaaaaaaa 120  
 aaaagggggg gggcgcccgga cgnagtgtga cgacgagatg tcgccgcgga cgaaacgccc 180  
 ggggggagtt ctccgggtgt gggggagacg ctccctcccc gctggtggta tgcgtttct 240  
 ccgacgagag cccccactt gtggtgctgc ggtttagtgc taccacacacc catcgggtgtt 300  
 tattttcgcc gttggtcccc cactttgtaa acaatatattg gagagggccc ccacgattat 360  
 cgctcgtaaa aaaaattctt gtggggggag tattaacaaa gaagtataac gccgggtaat 420  
 aagagaaaag tatacggggg aacatcttcc aaaaaaactt gaatatattt ggacaaatta 480  
 ttccccagg ggaaggagaa aaaaaaattt ggcccccttat tataaagccc cgggtttatg 540  
 gtaaagggga gcacacgga cggcgctgga acaataagaa ccattatttc aacacggtgc 600  
 gcaaacacaa ataaaaacac aacagcgggt gggggcacia acgagggcca caagagggtc 660  
 ccgggtgata acactgtgtc taccgcgcca caatcccaa ataatacaaa aacacagcgg 720  
 gc 722

<210> 28  
 <211> 1210  
 <212> DNA  
 <213> Homo sapien  
  
 <220>  
 <221> misc\_feature  
 <222> (631)..(631)  
 <223> a, c, g or t

<400> 28  
 aaccggcggt tttattacgg tcttgagtaa tttcccttgg ccaaattccc agttttgcca 60  
 ctgcgtggag ccagatcctg ggagctgtca gcaaggagca ggtaagtga cagttatgga 120  
 cagcaccttt ccaactgtgt gcttccgaac cctggctgtc acgagtgaat tgtaaagtca 180  
 cgggctctgt acagttttgc catttcactg ttctgcttta agcttagctt attagaactc 240  
 ttggtggagg gtgcgtacac accacttcca gaaaaggctt caccctcgtg ggaacgtcaa 300  
 cccagcgaga aaggagggga agcccccttct ccggggacct tatctgtgga cttaggaatg 360

## 16

atggtgttta ttgcaaatgc acaatctttt tccattgaa atgtcatcac actggaatt 420  
 gtactatatg taaaaaaaa aaaaaagtat agttttatat ttgaaatgta tgcaattat 480  
 ggccatatgg ctgattggaa tgtactactg taatataaaa agtcactgta ttgcaataa 540  
 attcttttct attaaaattg aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa 600  
 aaaaaaaaa aagggggggg gcgcccgcgc nactgtctac acgagatgtc gccgcggacg 660  
 aaacgcccgg ggggagttct ccgggtgtgg gggagacgct cctccccgc tggtggtatg 720  
 tcgtttctcc gacgagagcc cccacttgt ggtgtgcgg tttagttcta cccacacca 780  
 tcggtgttta ttttcgcgt tggccccca ctttgtaaac aatatttga gagggccccc 840  
 acgattatcg ctcgtaaaaa aaattcttgt ggggggagta ttaacaaaga agtataacgc 900  
 cgggtaataa gagaaaagta tacgggggaa catcttccaa aaaaacttga atatatttg 960  
 acaaattatt ccccagggg agggagaaaa aaaaatttgg ccccttatta taaagccccg 1020  
 ggtttatggt aaaggggagc acacgcgacg gcgctggaac aataagaacc attatttcaa 1080  
 cacggtgcgc aaacacaaat aaaaacacaa cagcgggtgg gggcacaac gagggccaca 1140  
 agagggtccc ggggtataac actgtgtcta ccgcgccaca atccccaat aatacaaaaa 1200  
 cacagcgggc 1210

<210> 29  
 <211> 247  
 <212> DNA  
 <213> Homo sapien

<400> 29  
 aaaaaaagg tagatttcca gataatttta cctggtccag caccgggaca cacctcccta 60  
 aatgcctgtg taataatatt tggaatctgg atcctgcatt tctccctcaa tttatgtact 120  
 ggacaactaa acttattatt tcatctaaaa aaattcaaaa acaacaaaca aaaaaaaaa 180  
 cgcgggggaa accaggcaca aaggggtccc ggtaaaatgg ttccgacaac aaaaacaaa 240  
 caaccga 247

<210> 30  
 <211> 528  
 <212> DNA  
 <213> Homo sapien

<400> 30  
 gaaaaaaga aataatctta tctgcaaaca ttgctgaaac ctgtgtagtt tcttctttt 60  
 tctcttggtg ttggtatcaa ggaatttaaa ttttagatgg actgtgttta ttaaaattgg 120  
 tagactatgc taaacaaatt tacaattctt ttgcctagaa aaatggaact acttaagtct 180  
 tatataactg gaaaactttt acttttcgct taacattaat tggaattttg gtgacagtga 240  
 aaattatttt ttttcagggc ttgttaaaca actgttttaa aacagatgat gaccaaacc 300



i7

ctgctcaatg agaatagtat tgtatgtgaa actctaaaga agtcattatt catctcattt 360  
 tgcagatgga attaagaatg caaaaatagt ggacatgccca agtgaatgct gttaataata 420  
 tgtaaaatta tttgattaac atttataact taaaaaaaaa aaaaaaaaaa aaaaaaaaaa 480  
 aggaaaaaaa aaaaaaaag ggggggtgggg gcactccggg gaaatccc 528

<210> 31  
 <211> 890  
 <212> DNA  
 <213> Homo sapien

<400> 31  
 tcgagtcagg aaagacttcc taagaaagggt gacacgagct gagtccttga gaatggggag 60  
 gattttctaga gatggggact cagagaaaag atggccttgt ggtcaaggga gaaaaggag 120  
 ctttagcttt ggctgaggca gaagaggggtg cagagatgtc acaagacaat ctaaaacca 180  
 tagagaagac acagtttgtgt gtctccacac ctgccctctt ggagtttga tggcaaagac 240  
 atgcgaggtg gttttgagca cacctaagggt ccgtttcagg ggtcctgaat gaggtgattg 300  
 cgacaactca aagactaagt ttctaagatc ccaggcatgg agtaaaagcaa ttctatacac 360  
 aggatctcaa tctagtcac aaagacttct taatgataca ggggctcaga gacatgggtt 420  
 cccctaaaca cgtcagcttg gattcatact ggccccatat tttccagtgt gccatgttgt 480  
 tatcctttat gaccctcgtc accatgccca cgtcccactc caaaataaaa atcaaagcaa 540  
 aacatataaa atatagtgc tgcaaatact tttaaagcac ttactatgca tcaggcttat 600  
 tatatccttt ttatactact acaggtctta caattttgct gtattatctc cattttgcta 660  
 gtaaggatat tgagatgcag agattaagca gttcgttcaa ggtcaccaag gcaggcaggt 720  
 gcaagggctc atgcctgtaa ttcccagcac tttgcggagg cccaaggtgg gttgggatgg 780  
 gtttgagacc caggagtcca aaaaccagcc tggcaaacat gggcaaacc atttctacta 840  
 aaatcctgat cctcaggccg atcaggaaaa gtggtcaact ccaactgcga 890

<210> 32  
 <211> 387  
 <212> DNA  
 <213> Homo sapien

<400> 32  
 catgcacacc aatccgagct gggctcgggc gccctggtga ggacaccaag cagccacgtt 60  
 gcctgtgctc cagcagctcc gaggtctctt cctggaagtc tgttggtgt catcctgcag 120  
 cccagagcca gggaaatggc agtggggagg gggcttcctg gggtgacagc aaagctctgt 180  
 gtccacaggc aggcaggacg catgctgcag ccctgtgggg tgggcacggg ggaagccttc 240  
 ctctgtgtgg cagaaaatgt gtctcagatc tctgggaact gggacaggaa agttcccaga 300

18

ggggcatgta tggggaggct acagaaagtg tccccccatt tcatgtttgt gatagcagct 360  
caggacagac aaacaccaag aggggtgg 387

<210> 33  
<211> 895  
<212> DNA  
<213> Homo sapien

<400> 33  
cttgactctt cagggtcttt gagaatcttg cagttggttt tcggtcacag ttgctttgca 60  
aaaactgaac tgctgaacag agtggcctga ctctctttac cctgtcccc tctccccagc 120  
ctggaatggg cctggctgcc caccgcacac gtggcaaggg cccctccttg tgccctgggg 180  
ctcctgagca gctttcctag gaggaagaac ctgcaccccc cagctatatc tttatgggat 240  
cctggcctgg actgaggaca aagccagggg ccacggggta cccaagctg cccattttcc 300  
tggaagggc acagtggccc tgaccggagc tgctattttc ggctggggtt ggtcagtcct 360  
gccctccttg ccgtggctgc tgcagcaca tgctattcat gtcgtaacca ttcgtggggc 420  
tccttctgc gccagcgtgg cggggctgag gccatgcaca ccaatccgag ctgggctcgg 480  
gcgcctgct gaggacacca agcagccacg ttgcctgtgc tccagcagct ccgaggtctc 540  
ttcttgaag tctgttgggt gtcctcctgc agctcagagc cagggaatg gcagtgggga 600  
gggggcttcc tggggtgaca gaaagctct gtgtccacag gcaggcagga cgcagctgc 660  
agccctgtgg ggtgggcacg gtggaagcct tcctctgtgt ggagaaaat gtgtctcaga 720  
tctctgggaa ctgggacagg aaagtccca gaggggcatg tatggggagg ctacagaaag 780  
tgtccccca tttcatgttt gtgatagcag ctgaggacag acaaacacca agagggtggc 840  
cttgggcagc agccagtga gagaggcaag atgggggtta gcttcgcaca ttgag 895

<210> 34  
<211> 502  
<212> DNA  
<213> Homo sapien

<400> 34  
aactctttta gacctagcat cagatgtctc ccttagtgag gtggagcccg gcaatggact 60  
gtctgctgct gcttctatgc ttattaagca ttccacctc atcataaaat atgtagcaat 120  
gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 180  
ctctctctct ccctcttttt ttttttttta ttctcccagc ggcaccccgc ggggggggga 240  
gggggatagg gggacacggc gggagggagc gaggcgagag cgcgcgaggc ggtagcagac 300  
acaatacaaa aggtgggtgga gaaccacggc gcgcaacaaa gtaggacgcc ggggaggaaa 360  
atgacgtcca cgcagcgcca caggcccacc cagctagcgc acgacgaaca cgacgagaga 420  
caacagccgc gagcgcagac cgcacacgaa cgaaccacaa aaacacagaa acacaacacc 480

ggagcggcgg cgcaagcgac ga 502

<210> 35  
 <211> 645  
 <212> DNA  
 <213> Homo sapien

<400> 35  
 actccagcct gggcgacaga gtgagattcc atctcaaaaa aaaaaaaagt ctgacataaa 60  
 accttggcaa gcagggtgctc atgggaattt ccaggggctc atataatttg gttggtgcaa 120  
 tgcctgtgga gtttggcatg cacttatatt ccctccatca aaaataacca caacataaag 180  
 agggtaaaagt tcaaagatca tctggctctg gatactacaa caaatagata actcttcttg 240  
 gatatatctt ttgggttaga aggagtgcaa ggaggagaa agtgtctagg tgatgagcca 300  
 agaaccattt aatcccatc aaacagccca ggtttctgc tgtcactgct gacttgacat 360  
 gggtaagaag gcccttgatc agctcaggat ccttagaagg cttccatcac agggttggcc 420  
 tgtaaagggg tgtatactac acaccaggat agatctcaca caacagcaac gagagaaaac 480  
 cagtcaggcc caaagtctgt caccttgttg ctcaatcttc accatctctg tatcatgtag 540  
 acagtccaat tggaggtatc aggccattcc ccaaatacta ctattttaag ctgggtatca 600  
 tggcatagct gtccttgtga atgatcggtc aatccccata cacca 645

<210> 36  
 <211> 173  
 <212> DNA  
 <213> Homo sapien

<400> 36  
 cacattcact ttttaatttt cgagtatcaa ccattaataaa aaattccttt catacataaa 60  
 tacatgttga tttccaggat ttcaaaccat ctacttaagt tttatgcctt aataggagtt 120  
 gctattcagg actttaaaaa gatcttcgaa ccttcacaat agctcaatat tca 173

<210> 37  
 <211> 858  
 <212> DNA  
 <213> Homo sapien

<400> 37  
 gcgagtattt ttttttgaga cagagtctcg ctctgtcgcc caggctggag tgcagtgacg 60  
 ccattctcggc tcattgcaac ctctgtctct caggttcaag taattctcct gcctcagcct 120  
 cctgagtagc tgggattata ggcatgcgcc accacaccct gctaattttt tgtattttta 180  
 gttagatagg gctttcactg tgtagccag gatgggtctca atctcctgaa ctcatgatcc 240  
 accgccttg gcctgcaaaa gtgctgggat tacaggtgca cgccgccacc ctccgctaata 300  
 tcacattcac tttttaattt tcgagtatca atcattaaaa aaaattcctt tcatacataa 360

20

atacatgttg atttccagga tttcaaacca tctacttaag ttttatgcct taataggagt 420  
 tgctattcag gactttaaaa agattttcga acottcacia tagctcaata ttcaaagctt 480  
 atttccctaag gctaaacagc acaaataatt tacccatgtg gcaattaaga tactgaaaag 540  
 taccaaatct tgacaaaacc tctgtgaac tctatttggc actcaaattg gcttcaggtc 600  
 taattttatg tgtttgaaa ttttggattt gattccaccc atatttggct tctgtcaca 660  
 attcattttt cacaaacaca gtaattctca ttttattttt ttattaaat tttttctttt 720  
 aaaaaagtag agacgagatc tctactaaagc gtccaggctg gcttcaaact ccttggcctt 780  
 ccagtgatct ttctacctca gcctccctag cgtgtttggg actgcgcag agtcacggca 840  
 atgggccag ccatcact 858

<210> 38  
 <211> 1314  
 <212> DNA  
 <213> Homo sapien

<400> 38  
 acaaataaaa cagatgttcg ctcagatgta tgacaagagg ctgtgcacag acaggatgga 60  
 acgagctctc gcgtatgagg tggaagcacc catcagaggg ccgaccaggc gccgcaggtc 120  
 ggcacacaca accaaacacg aagcgtcaga ccgtcagcca tatgaaccaa cgagagtcag 180  
 cgcaacgata gatcgaaccg gagcgtaaac accggacagc gaagatgacc acgagcacia 240  
 aagggaaaca caacacatca ccaaggctc gcataccacc gccacccgac ccaacacgag 300  
 gcactacact ccacaaccac accccgaccc taatagcgca cagccactca ctgcgaacc 360  
 acgcaggacg aacagggcac acacccaacc acatcgcaa agcatgacca cacacgaacc 420  
 ccaccaagg cacaacacg ctaccacgcc cgcgcgcaca caccgcccc accacgagcc 480  
 ccacaccccc cccacacaa cccccacctc gccaaaccaca accccggcaa caacccacg 540  
 cacacacacc accaccccca ccacagcaca aaccaggaga gaccggacag cagagaaaac 600  
 gacacaacga gggggaaaag aggacaacga cgcggagggg cgcagaaaaga gggggccgat 660  
 caccaccacg gcgagcggcg cagagagcag agggggccta gcacgtcgcg cgcggtggcc 720  
 gcccgcgaat acgacgcgac acgccacgaa cgacccaaca caccagcgca ccgcacagca 780  
 gcaaaggcga acagcgcgag accagagagg aacagcggac agacacaccg atgcgagagg 840  
 ccacgaccag cgacgccgca caacggggga tgacacaagg caggcgacgc agcgagcgca 900  
 acccacagga agggaaagaga agagggggaa gaagaacgcg aaggcgagac cagcagccaa 960  
 caggggagct aacggccac aggcgcggc agcacacgag taaggtaagg cacagggaac 1020  
 ggatacagca caggaggagg gcaaggaccg gcacgaagac acagggaaac agaggcgcg 1080  
 atggcccaca gaaacggcag aaaagaacgc gggaacggac cagaacacgg ggcagcaagc 1140

21

gcacggagga gccagaggca gcacaaggga accgaaggac gaaggggacc cacaagcaac 1200  
acgggacgca ccacaggagc gaaccaagca caaggaagca cagaggggga acacaaacga 1260  
cgaagcgacg cagccgacgc agaacgatga aacgacagag cgacaaggcc acac 1314

<210> 39  
<211> 418  
<212> DNA  
<213> Homo sapien

<400> 39  
tggtcgcggc cgaggtcttt tttttttttt tttttttttt ttttttattt tggaatgttt 60  
tttataaatt ttattttttc aaaataatga ctttagtaaa aatttaacat acccgttttt 120  
ggaatcccc ctttcaaatg aggcttcccc agtaatgagg gggattaatc cagaccctag 180  
tgtttgtggc atttgtgact ttactcctc aaaagtgagc atacacgtgc ctcacagtga 240  
attatccag aagaacttca ttactctttt tatatttttt ctccgtggaa aatttaacaa 300  
aagaaaaagc ttggcgggct acactcagtg gctcataggc gtggatctcc gtgggtgtga 360  
caattgtgta tactccccgt ctcacacttc tccacacaaac tattaccgga ccaacaca 418

<210> 40  
<211> 672  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (255)..(412)  
<223> a, c, g or t

<400> 40  
gccgccccgg caggtacgcg tgtatgtacc tgcgcgcttg cggggacgtg cttgtggcgg 60  
gcggcgagag ggatgggctg ggctaatatg aaagctgcat ctttactagt tagctaccat 120  
gcgtcattat ttatcaaaag atatatgctg cttaaacaca aatacgtttt aaaatatatt 180  
ttaggcagta gggttttggg tttttttttt tgcaagttct ttgggtgagt aaatttagtg 240  
ataaatgatt ttttntntnn nntntntntnn nntntntntnn nntntntntnn nntntntntnn 300  
nntntntntnn nntntntntnn nntntntntnn nntntntntnn nntntntntnn nntntntntnn 360  
nntntntntnn nntntntntnn nntntntntnn nntntntntnn nntntntntnn nnaatgcaaa 420  
atgccatcga cgcagaaaaa caaagaatca gcttaagttc cagaaaaaag aaaaaccaac 480  
caaatgaacc ataagacaac aacaacaaca acaaaaaagg gctttgggga ttcgggatga 540  
tattgcatca aatccataca ttacctgag agaagagcga ttcttcaaca ttgagccttc 600  
caatcatgat ttccacttca tttaggcttc tgtaagggcc tcacataatg gatttgtgca 660

tgcgcaagtt cc

672

&lt;210&gt; 41

&lt;211&gt; 687

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 41

gcgtggtcgc tggcgagggt tttttttttt ttttttttgg aaaagggtaa atttataagg 60

gaccccgtaa aattttaaaa aaaaacaatt acaaagacaa ataaaaacat ctgaaattaa 120

tttggcataa cagaacacaa aacttggttc aacaactcca cagagttaat tactcaatat 180

aatctcctc catgtgggaa caaaatttca tttgtgcctt catagtagaa caagagtctc 240

atctcgcat atacccttcg agtctcttat acaattctca cagaaacgtg ataaaattag 300

cctcaaattg gacaaggaga aagagatggg agacccttg tagcatctca cgtgtcaggc 360

ctccggagaa gggctctgta tagggataac tccctataga ctcttgggtcc aagaagaaga 420

ccccaggga attgggtcttg gccattctc aaaggctctc ctcatagggt ctccattggg 480

caaacccagt gcccgcaaca cacggaggca gcctcataaa ctcatataatt aatggggcac 540

tttatattaa aagttcagcg ttattcctcg tgattaataa aatctactgt gtgggttcaaa 600

aaaggctggg cgataatcat ggggtcaaagg ctgtttccct ggggtgaaat gggttatccg 660

ctcaaattcc acaaattgca aaaaaaa 687

&lt;210&gt; 42

&lt;211&gt; 63

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 42

ccccttgtag cttgtggacc atacaaaaac actgtggcca gatttggctg ctgggttgta 60

att 63

&lt;210&gt; 43

&lt;211&gt; 470

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 43

gccggggcag gtccctccc tttttttttt tttttttttt tttttttttt tttggtaggg 60

gaaaattttt ttttttaagg gggtttccca aaaaaaaat ttttcaggga atataaaata 120

aaaatctatt taaaaattta tcccagggtta ttacatttcc cctccctccc caaaaggcta 180

catttgggag tacaaaaaac atccagtgtc ttaaaacacc tggatctctg gttgcggcga 240

cgttaaagag gaggcaagat agctggcgct ctcacaagca cactctaggg ggtggtcccc 300

cttacgggag ggggagggat atgcgcccc ctattacact cttgggtgca agggacaaga 360

23

ataaaaaagtc gtgggcggta ccccggggcg catcagcgtg gtgcccgggg gggaaaatgg 420  
ggaatccggc ccacaatccc ccacaaacta tccccgccca acgaacacgc 470

<210> 44  
<211> 713  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (45)..(463)  
<223> a, c, g or t

<400> 44  
cgaggtaccg cgccagccca ggagaacccg gaagccagca gctcnnnnnn nnnnnnnnnn 60  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnaaggccc tcgcccgcgc 480  
ccgcgccccc cccgcccccc ccgaagacag cgaggtcagc agagaccgaa gaatcaacca 540  
ccaacgacca gcgaaagagg cacatcacia aaggaaagac agcatcgagc acacaacgca 600  
ggctacaaac ataagcgcga cgaaccatat agcgactgga gtacaggcaa aacaagacat 660  
tatatgactg gcacgaccgg tgcgcacacc gctgatagca gacgacacag aag 713

<210> 45  
<211> 488  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (254)..(365)  
<223> a, c, g or t

<400> 45  
acttcagtca atgtcgtgtt agagtggagg aaatatagta acacttcatt ctatgaatag 60  
gccaccatta atgtaagcat tcctctgttg aaagacattt ggattctttc ctgttttttc 120  
tgtttatgta tgtatgtatt atatttttta ccttgaggca ttcttggaac ttcttcttgc 180  
acacttgagc acttaggaca gttttgcaaa cttctctggt gttaccagtt acttaggcat 240  
ttatgtaaaa atannnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300

```

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnaaaaa aaataaaaaa aaaaaaaagt tgggggttaa cagtgggcca ttacggtgtt 420
cccggtgtggt aaaatggttt attccgcccc aaattttccc cacaattttc ccaccaacaa 480
tacaagag                                     488

```

```

<210> 46
<211> 487
<212> DNA
<213> Homo sapien

```

```

<400> 46
ccccagtgtg atggatcgag cggcgccccg gcagggtgct gggccagacg cttcactctt 60
ctgtgaaagg aaaacggagg gtagggattt ttaaacctac atgtttccca gggcctgggg 120
caagtcttga gtagactgtt gcagtaaacc gactcaaagg cctatcacct ttcttgtgag 180
gctcaaggtc taatcattaa ttgacatgaa aaccacagga gagaagcaaa cccttctgtg 240
ctgggatctg tgccccagtg ctccatgttc cctgataggc ggctaattga attcataaaa 300
taaatgacat gcctcttcct aaaaaagaaa aaaaaaaaaa acaaaaaaaaaa aaagagagct 360
tgggggttac tccaatgtgg ctcatagcgg tgttccccgt gggttgaaaa tgtgggtttc 420
tccggcctcc acaattctcc ccacacactt ttgcacccc aaggggtcgg agcggaggaa 480
gacaagc                                     487

```

```

<210> 47
<211> 667
<212> DNA
<213> Homo sapien

```

```

<400> 47
gcgtgggtgc ggcccagggt ccataaccct gccctcatcc cagatctgtg cagatgaaag 60
agagggaggg agaggggaaag agagagatgc tttgggggtg atttggccag aggccaccag 120
gctggatccc atgaagaaat ctgggtgaga gggctcttaa gtcataaact gagatccagt 180
tgccagggtg ctgcatagtt gccaacagtg taatgtgtca ccttttgatc ttcacagaa 240
atctcagcct ggtggccaac tgccaaata cactgcagag catgtctgtc tgtctgtctg 300
tctgtgtctc tctgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg 360
tgtctcctca ctctttcatc ctatcattac atagtagtat aataataaat attagagaga 420
tacacagaaa atatatagag aagataacag tgttctctat aaaaaaaaaa cagctgcctt 480
ctctgcatag cttctaacaa ctgagcaact ctgcagaaa agagcacaaa acgggagaaa 540
caagaacaaa acgggagaca agactagaga aaacacagga cagcggacaa aaccacgtga 600
gggagcaaca ccagaggggc gaaccacatt accccacaca cgtgaaaaag cgagaccagg 660

```



gggggaga

667

&lt;210&gt; 48

&lt;211&gt; 1677

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 48

gagttgcggc gtgccaaggc ccacgagggc ttgggcttca gcatccgtgg gggctcggag 60  
 cacggcgtgg gcatctacgt gtctctggtg gaaccaggct ctctagctga gaaggaagga 120  
 ctgcggttcg gggaccagat tctgcgcgtc aacgacaaat ccctggcccc ggtgacccac 180  
 gcggaggccg tcaaggctct gaagggtcc aagaagctgg tgctgtctgt gtactcagca 240  
 gggcgcatcc ctgggggcta cgtcaccaac cacatctaca cctgggtgga cccgcagggc 300  
 cgcagcatct cccaccctc gggcctgccc cagccccacg gtggtgccct gaggcagcag 360  
 gaggtgacc ggaggagcac cctgcacctc ctgcaaggag gggatgagaa aaaggtgaac 420  
 ctggtgctgg gggacggccg gtccctgggc ctacgatcc gtgggggagc tgagtacggc 480  
 cttggcattt acatcactgg cgtggacca ggctctgaag cagaaggcag cgggctcaag 540  
 gttggggacc agattctaga agtgaatggg cggagcttcc tcaacatcct acacgacgag 600  
 gctgtcaggc tgcttaagtc atctcggcac ctcatcctga cagtgaagga cgtcgggagg 660  
 ctgccccatg cccgcaccac tgtggacgag accaagtga tgcacagttc ccggatcagg 720  
 gagaccatgg cgaactcggc aggggtctggc cactctgctc gctccaatct ccagacccca 780  
 gggccatttc tgaaagccag tgatagctgc ctcccatccc tccaccgccc tggctctcct 840  
 ctacgctgc agtccccaca ccagggccct ccattggcag gacatgacct gggcacatcc 900  
 ctctctctc ttggcctcag ttcccccag gaaagctgaa atacaccatc caactgtctc 960  
 attctttatt tgtcccaaa ttacttaact cattctatag accttagttg cttcatccaa 1020  
 aaagtgggga ccataaccct gccctcatcc cagatctgtg cagatgaaag agaggaggag 1080  
 agagggaaa agagagatgc tttggggtgt atttggccag aggccaccag gctggatccc 1140  
 atgaagaaat ctgggtgaga ggggtcttaa gtcataaaact gagatccagt tgccagggtg 1200  
 ctgcatagtt gccaacagtg taatgtgtca ccttttgatc ttcacagaa atctcaggct 1260  
 ggtggccacc tggccaaata cactgcagag catgtctgtc tgtctgtctg tctgtgtctc 1320  
 tctgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtctcctca 1380  
 ctctttctac ctatcattac atagtagtat aataataaat attagagaga tacacagaaa 1440  
 atatatagag aagataacag tgttctctat aaaaaaaaa cagctgccct ctctgcatag 1500  
 cttctaaca ctcagcaact ctgcagaaa agagcacaaa acgggagaaa caagaacaaa 1560  
 acgggagaca agactagaga aaacacagga cagcggacaa aaccacgtga gggagcaaca 1620

26

ccagaggggc gaaccacatt accccacaca cgtgaaaaag cgagaccagg ggggaga 1677

<210> 49  
 <211> 802  
 <212> DNA  
 <213> Homo sapien

<400> 49  
 aaaaaaaaa aatTTTTTTT caacattaaa tTTtaattga aaacatgaat atggctgggt 60  
 gctgggtggct cacacttggT aatcccagca actctgagaa gaacagaagg gtggggTgga 120  
 atcccaagca ctttTgtgaag ttcaaacagt tGtaaaacca gccgtgggtt aacacgggac 180  
 tccatctcta caaaaaaaaa aaaaaaaaaa aatggggggT ggggggcatg tggcgccgtg 240  
 ttacccccag agttaacccc taaaagctct ggggtggggg agaggaaCTg gctgggagcc 300  
 cccgggaagt tgggaaacct gcgagtaagc cttaaggaca ctcccgcgga gtggccact 360  
 cccaaggcgg gaaagtggag gagaaccaa aacttgtggc cctcaaaaa cacagaaaa 420  
 acaattacat tcccagagtt cccgggacat ctctcttaa cctccagaga ggccccaaaa 480  
 ggagaaccgc gtggaaaacc gagggaaacc cctctcaaac tgaccgggtg gaaccacagg 540  
 cgcgacacac ggcgaaacct ggggggggaa ccccaaaaac acagatcccc caaataaaaa 600  
 ggggggcaca acgcggggct ccccagaga caccaccggc gctgcgggac ccccgggcgc 660  
 cgaggaaac aagggcgaac acgcattggc ggcaaaaggc cgtgggcggt aacccccacg 720  
 ggggcaaaaa ccgctggatg cccgggctgt aacacagggg gataatcccg gccacaagg 780  
 ccccaatac cagcaccac aa 802

<210> 50  
 <211> 918  
 <212> DNA  
 <213> Homo sapien

<400> 50  
 gaagaacccc gggatgttag atatattggc atgctgatct agatgcatgc tcgagccggc 60  
 gccaatgtga tggatgcgtg gtcgggcgga ggtaccaaaa tacagaagct gattccaaaa 120  
 tctatgctcc ataaccatcc gagactgcc aggctgcaat ccatggagac agcgagaaac 180  
 atgacaacaa acaatacat tggccgagtc tgaaatctga ctctggttcc taattctacc 240  
 actaaacttt ttataatttc tgattataaa aataatgtga aaataacata gcaattaaca 300  
 tctattgatc acttgggact aagcatctgc cagagatcat ttaattctca cctacaagt 360  
 agatactatt ttctggggT gaagggattg gtctaaggTc atagagctat catgtgtaag 420  
 aggcaagata agattcagac tcaaaaggcc agaggatcag agttacactg ctttctgca 480  
 cagaattact actgattgtt gccccggtta cataggactg ctgagaaaat ggcacacaga 540  
 cttatttctt cggagaaacg tcaaatgttt catatgattc attattctta tttttacttt 600

27

```

tgaatttggg gttcattgtt taattataaa agatggctct tttactagca aaaaaaaaaa 660
acaaaaaaaa aaggcctggg gggtagcctc gggttcataa gcgggtcccc ctgggtggac 720
attggttatg ccgcgccaca attccccaca atttacgact acacaacgta ctagcaagca 780
ccagactacg aactaaaca tcacacacaa cagtcaaaaa acagccaccc gaacacagca 840
aaacacaaaa acttcaacac atcacacaac agaacgacaa agagaatcaa caacaaagc 900
ggacaacaac cacacacg 918

```

```

<210> 51
<211> 985
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (856)..(856)
<223> a, c, g or t

```

```

<400> 51
taggcgaatg gtccattaat ggcattgctc gagcggcgca ggtgatggat gccgccgggc 60
aggctctctc ctctctctgt gtgtgtgtgt gtgtgtgtgt tatatgtgtg tgtgtatgga 120
gggtaggtga aaggggatga ggaatttatt tctgtcttcc tgggaaggata gattcttctt 180
ttttgaatta gcctcattaa acttttaagt aatgactcct gaaaaaggac aaagggataa 240
ggctcttttc caaagagtta tctttgtgtg ccagcaatca gtcattactc tctaccatg 300
ccatgtgaca caggatgtgg tctgatattt agtctaaata catgcttcac ttttttctg 360
ctacagagaa ggcaattata atgctccttt tgttatgcaa ataacttctc agaaaagtgc 420
cctctctcct ccttaaaaac tagatttact cagactaggg tgaaaaataa aagtcaatcc 480
tggcatttaa gtggtttctg gccctcagaa gccatcttag tagaaggga tgaatatgtt 540
tcagtggctt cctacttctg gaatatgagc agggtcagtc tacagcagag tcagaagggc 600
tgtccctcca gggatccagg aaggctgtaa cctcagtgtg taacccagc ctttggggga 660
acaaagttag acacttctga agtgttctgt atttcatttc ttgggaccct aaccccataa 720
actataataa aatggggtaa gtggaatgag tgtaataaat caacctttt cactcacata 780
acgttagctg ttataattat tcttttatgt aacaaatgcc taagttaggt atgggttttc 840
tagaaaattc agggangggg ggggaaatac ttaaacaggg ctttcaaac caagcaatat 900
gttgtttgtt tgctccata cgaagcttg gtttccaaa ggggggggcc caggggaaag 960
agctttttta aggaaacaaa aacac 985

```

```

<210> 52
<211> 669

```

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 52

```

ccgcccgggc aggtactagt agtcagggcc ctcagtctca ctttgcccc tgacttgatc      60
gagttcactt ccttctcaat aaacatggca ttaggccaga caatatttaa gcagagtatg      120
gtggaaatgt gggcaggtct gaggggtggg aaaataaaag gataaaatac ccctgaggga      180
ttagatatat ttaaaatcac aaaggtatta taccacagat ctataacttt actaaaatat      240
aaaaatgaat gaaaatatat ttggtattat tttatcttag ccctgtaaga gaagctaatt      300
ttctcttggt gctcttcagt ttttagtaag agaagtgcaa gcaacttttt cttatgggcc      360
gggatgaaaa atagccttat gaactccag gaggagtttt ttcttaaggg gatacatatc      420
atttaacca cagaagagag gtaagtaaag ggtgagtaac ctagattgtc tagaaaaagg      480
tggtattaga gagacccttt atgtattcta gagttgcaga gttgtgtagg aaataacact      540
gccacctata cctatggaca tgattagaaa gaaacaatgg gaggcagttc tgtaacagtg      600
gaatcatttg actcaaagtt gggtaatcag gtcatagctg tttctgtgtg aatgttatcg      660
tcacatcaa                                     669

```

&lt;210&gt; 53

&lt;211&gt; 837

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 53

```

aaggatgata tctatagggc gaatggctct tagatgctgc tcgagcggcc gcagtgtgat      60
ggatgcgccc gggcaggtac agcttttttt tttttttttt ttttgggaaa tggaaatctg      120
ctctgtcacc caggttggat taaagtggcg caaccttggg tcaccgaaac ctctgcctcc      180
tgggtggtgt tcaaaaatat tctcctatcc tectgtgtcc ttfcagcttc tcccaagtta      240
gctgtggggc ttacaggact tgccaccacc gccaccacgc ttaattttgt gcacgttttt      300
agtaaagcac gggggttctc acttaatttg tttggcccag gcgtgggtctc tcgactcctc      360
cgtgaaccgc aggtgactcc ctccgtgccc tcgcgcctcc tgaaaatgtg gctgggtgat      420
taaacatggt tggtagacca acctattggt ccagccacaa aaaatattat tttcttaatg      480
tcaatgtttt tggagtcttc aacaccttat taattctttt ctacagtggg ctattatact      540
aatattatcc cccaatattg ggatattatt attggagatt gttgttatcc acaaatatgg      600
agaatatgaa tatgggcgaa atatcgctaa aaagaaatct tcagtattcc ttattattca      660
aatgttatcc acaaatatta ttctcacaaa atatttcttg aactctataa acaaaaatat      720
aaaaaaaaa aaaaaaaaag gcttgggggt actcttgggc caaaactggg cccctgggtc      780
gaaattggtt cccgtcccaa tcccacctcc tccaacaaaa aggaaaaaaa gaaaaaa      837

```

<210> 54  
 <211> 718  
 <212> DNA  
 <213> Homo sapien

<400> 54  
 gggaaaacaa tgaaaagaaa tgcacgtag ttttcaatcc agatttaaga agtaacaaca 60  
 atcttttttg ttcgtgcgtt gtaaaggaca aggtctcact cgtgttggcc cagtgcctgg 120  
 gaagtcgccg atggatgcaa atcaatgaat cttacttgca ttccttgta tcctatcctg 180  
 gggcatcagt gtgatcctgc ccaatctcga gccatcccg agggaagctg ggtactcaac 240  
 taggtcgtag cactacgcta agccatcgct ggcataattg ttcactattt gccataagga 300  
 cagggttgtt cgccaatgtc tggcccaggc tgaagtcatt ggaatctacc atgtggcact 360  
 cgaatggtcg agttcataac cctaacgctg tggagcgtcc acaagagtgc tgggtgattta 420  
 cgaacgggtt tacatgtcac tagcacatca gcacaaacag atctttaatt ctacgaggat 480  
 gataggatct ctgtatatag aacacatcct aaggattgct atcaggataa aaattattag 540  
 actatgaggt tggagacaag ggtcgcagaa taaatgtgta tttctacaca cgagcaatga 600  
 acaatctgaa catgaaataa taaaacaatt ataacagca ttaaagacag cttggcgtag 660  
 catgtcatag ctgttcctgt gtgaaatgta ttccgtcaca ttcacacact agagcagg 718

<210> 55  
 <211> 913  
 <212> DNA  
 <213> Homo sapien

<400> 55  
 cgagcggcgc ccgggcaggt actgacctga aaacttgta caagaatgaa caccaacaag 60  
 tgctccctgg gactgtagtg accctttctt gccatcccca tccccgtgaa gtctgaacct 120  
 tgaggagac aacgagtcgg agggagttag ctaggcgat gcaaactata ctagaatgga 180  
 gtgccttgga gggtcataat atgttaggaa tggatagata gaggaaatgg aggatgataa 240  
 agatggcagc atacataggg gtacatacag tcaagaaaga gtggaaaaat agggaatgac 300  
 atgaggaagg gatgaaagtg gtagagtgcc attgtaattt gcatgagtaa tgctggaagg 360  
 ataggtcgcg gagcggtagg acatgatgaa gtggtaggcg catgtgaaga gggaaacgcg 420  
 aagatgatgc cttcaggagc gtttcgtgac tcgtctaccg tgggggggta tatcaggggg 480  
 gcatagcatt aaaatagtaa catccctatc gtgaatttac tatcttggt tactaggagt 540  
 catggtttat atggcgctcc atgcaaagaa gtgctacggc tcagggcact aacactaagg 600  
 tgcaattttc gctacctcgt ttctcgtgcg acgttggtgca gtggctggtt actgtgcgta 660  
 ttaagaggcc acctatttgc acagagagtg agagcaattc aacacataag ggataaatgg 720  
 ggctgggcaa ggctagttag tagcccaagc gtggccacgg gtgttgacct gttagggcct 780

gacagcattt gacttttagc caacaagag ttccggctgt gggaaatctg ttagtcaaac 840  
 attcgcctaa cttccaggca aatcttcggt agctagcttg ggaatcagtg ctgtgtccgc 900  
 gcatgttcct cct 913

<210> 56  
 <211> 1203  
 <212> DNA  
 <213> Homo sapien

<400> 56  
 ccctcaaaac tgactctggt ccacaataag ggcttttagt ccctggccgg ggacatcttg 60  
 atcaagttag aaggccgaag atcagtaaga tggatttgct gaataggtag atattctggg 120  
 tgtgtgtgtg tgtatgtatg tgtgggtgtg tgtgtgtgtg tgtgtgtgtg ttgggtgtga 180  
 taaaaacggg gagcaatgct aagatttctc atgagggttg atttacttta aacagtttat 240  
 accctectac cctaaccatc cattcacacc atgacacctg tgccttctc cctctaggga 300  
 aacggcaaca agcctcccag tactgacctg aaaacttggt acaagaagaa caccaacaag 360  
 tgctccctgg gctgaggacc ctttcttgcc tccccacccc ggaagctgaa cctgagggag 420  
 acaacggcag agggagttag ctaggcgat gcaaaactata ctagaatgga gtgccttgga 480  
 gggtcataat atgttaggaa tggatagata gaggaaatgg aggatgataa agatggcagc 540  
 atacataggg gtacatacag tcaagaaaga gtggaaaaat agggaaatgac atgaggaaag 600  
 gatgaaagtg gtagagtgcc attgtaattt gcatgagtaa tgctggaaag ataggctcgcg 660  
 gagcggtagg acatgatgaa gtggtaggcg catgtgaaga gggaaacgcgc aagatgatgc 720  
 cttcaggagc gtttcgtgac tcgtctaccg tggggggtta tatcaggggg gcatagcatt 780  
 aaaatagtaa catccctatc gtgaatttac tatctttggt tactaggagt catggtttat 840  
 atggcgctcc atgcaaagaa gtgctacggc tcagggcact aacactaagg tgcaatttct 900  
 gctacctcgt ttctcgtgcg acgttggtgca gtggtcgttt actgtgcgta ttaagaggcc 960  
 acctatttgc acagagagtg agagcaattc aacacataag ggataaatgg ggctgggcaa 1020  
 ggctagttag tagcccaagc gtggccacgg gtgttgacct gttagggcct gacagcattt 1080  
 gacttttagc caacaagag ttccggctgt gggaaatctg ttagtcaaac attcgcctaa 1140  
 cttccaggca aatcttcggt agctagcttg ggaatcagtg ctgtgtccgc gcatgttcct 1200  
 cct 1203

<210> 57  
 <211> 377  
 <212> DNA  
 <213> Homo sapien

<400> 57

31

cggcctcaca aagtgcctggg attacaggca tgagccactg caccagcct ggggaatctt 60  
 ttataatggg ttatgaagtt tacagacttc attcagattc cactaaattg gattttatga 120  
 gaattcagct gcagctgaca ttacctctg gtctaactct gaaaagaaaa attgtttccc 180  
 aaaaggattt gtggtatag tagtattaag ggtggggaag ggctatttaa ttaggtaag 240  
 ataaagaact ggttttaaga actttacata gtgattacat agaaatggat gtgggtagtt 300  
 acaaagggtt cttatctatt cattcatgcc cacctgccc gccccctgct gattcagacc 360  
 agctttcact gccaaga 377

&lt;210&gt; 58

&lt;211&gt; 1527

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 58

ggaggcttat tcgccgagag ttttttccca ccttgaggga tgttttcgcc cggcctgttg 60  
 tccccctctgt ttgccaggt tatgaaggct gtgtgcccag agatgtgtgg gaagacccgg 120  
 gagccccctt tgggggcccgt cccctttatc tcggtttaat agggccccag ggagtgcgcg 180  
 gccttggttg cgcttttttag tgactcgtac cccctttttg aatcgccacg ccaaaacctg 240  
 tggagatgtt ttttccccgc gaaagactgt ggggacaagg caaatcgggt tgggggcccc 300  
 acagggcttg cacacaaatg gcttgggcgc cttcctggag acacatctgt gggggaacac 360  
 acgggtttga aagcagttgc aaaccaaggg aggattgtcc ccggggtttt ttgtgaggat 420  
 taggtgaacc cccccacgtg tgtgaaaagt ttttaagttcg tgagctgttc gaaccgcacc 480  
 gcttgatata ttttcttccc cgggggtgtg gaaggcccc cgggtgtgcaa cacactgggg 540  
 gggatatatag ccgtcccccc caggggcgtg ttttcgcgtt gtaaaacttt tcccgggggc 600  
 acccccccg ggggtgttta aactggagag ggagtttttt tttccgcgtt ggaaacattg 660  
 tcacacacac gttggaggcc tgttgtaacc ccggagggtt gtggattgta gacagatatt 720  
 gaagcgagga gatccacttc ttggttgaga agggccccac ctggagggtg aaatcttata 780  
 actcgggggt ttttctgga gaaaagaaaa gttcctcgag attcgcgccg cgggagagcc 840  
 ctctctaata tggtaatat cgtttggaga catctcacac agaaaaatgg ccccaaacac 900  
 gctctgagtg tggagaagtg atacattgag aagagagggt ctccaaggaa gaactctttt 960  
 gtggggccaa cgcgcacagt gttcacacac acaacatttc tgttctcttc tttgggagtt 1020  
 tgaccgcgag ttgaacgggc tcaccgcgag agggccaata tttttttaaa aaccacactc 1080  
 ttggcacaaa cacattgtgg gtcaccaatg cacaaattat ggtgggtcaa taatgaccac 1140  
 gactgcacat tccgggagaa caaggggtaa gcacaataac ttgctttgag agaatcacca 1200  
 ctttcgaact cggctcgtg agtctgaggt ttttagatgt ttaaaaaatt taatgtggag 1260

32

aattaaatta aaaggtatgt tggctatatt cgctaccaca ttccacattc ttttgagcct 1320  
 tatgtgaata ttttactgga aaataagact aataaattgt taacagtttt taaaaaaca 1380  
 acaaaaaaga acaaaaaaaa aaagaaaaaa caaacggcca caccgcaccc ccgggcaaac 1440  
 acggcccccg gggggccctcc ggccccctc gccccccccc gcaacttttg tccccccgccc 1500  
 ccaccccccc ccacttcccc cacacct 1527

<210> 59  
 <211> 532  
 <212> DNA  
 <213> Homo sapien

<400> 59  
 cgcccgggca ggtacgtaga tgccattgcc atagccatcg ttggattttc agtgaccatc 60  
 tccatggcca agacctgagc aaataaacat ggctaccagg ttgacggcaa tcaggagctc 120  
 attgccctgg gactgtgcaa ttccattggc tcaactcttc agacctttc aatttcatgc 180  
 tccttgctc gaagccttgt tcaggaggga accggtggga agacacagct tgcaggttgt 240  
 tggcctcatt aatgattctg ctggtcatat tagcaactgg attcctcttt gaatcattgg 300  
 cccaggttgg ggtggtggc catggtgatg tgtcaacctg aagggaatgt ttatgcgggt 360  
 ctcagatctc ccctttttct ggagaaccag caaaatagag ctgaccatct ggcttaccac 420  
 ttttgtgtcc tccttgctcc tgggattgga ctatggtttg atcactgctg tgatcattgc 480  
 tctgctgact gtgatttaca gaacacagag tccaagctac aaagtctctg ga 532

<210> 60  
 <211> 499  
 <212> DNA  
 <213> Homo sapien

<400> 60  
 tttttttttt tattcaaaag tggaatttat ttctgacagc tctgaaggct gagaagctca 60  
 aagttgaggg gctgcatctg gtgagggcct tcttctgggt gggaaactgt cagaatcctg 120  
 aggtgacagg gcatcacatg atgtgctggc tcagttctct tccccctgct tagaaagcca 180  
 ccagtccccc ttttgtgaca tcccatat caatcaaccc atgaatcctt gcgcgggtta 240  
 atctattaat gagggcagag ccctcattga ccaatcaccc cttagagagc cccacacctt 300  
 taatactgcc acattgagga ttgagtctag aggggaatgc taccattcca cccctgatcc 360  
 cccaaaatca tttccttctc acattcattc tactcccata gttccaaagt ctgaactaat 420  
 tccagcacia aattccagtt caaagtcag agcctcactg tgtgagcctg tgaacacaaa 480  
 acaagctctc ttcttccaa 499

<210> 61  
 <211> 544



33

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 61

|   |     |
|---|-----|
| tggtcgcggc gaggtacttc tggtccttcc accctagccc cacctatcct ctccccatcc   | 60  |
| aagagcaaac agctctgaac agtctggagt agctggagac actcctcatc ttggcactct   | 120 |
| ccttgccact tgccatctag cagagctgga tgcttccctt gagcgtctc tgctccatcc    | 180 |
| cccaggtatc taggtgcct cccatctccc ccaactggcat ttgaacttta agagcctggt   | 240 |
| ctttgtgctt ggaatccaat gcaaaggctt ccataaacta gcactccata aacaactttt   | 300 |
| gaacaaaaat tcaaatccc agtggttcag ttgcaccaag ttcaagacta agtatttcaa    | 360 |
| ataaaaaaaaa aacaaaaaaaa aacaaaaaag ggcttgggcy gaacctccat gggcatctag | 420 |
| ctggttcccc gtttgtgtgg tcattgggta tccggctcac atttcccaca cactttcccg   | 480 |
| gcccacacag cagatgtgag agagacaata tccgcgccga gacgcagcaa cacaccgcca   | 540 |
| cacg  | 544 |

&lt;210&gt; 62

&lt;211&gt; 589

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 62

|  |     |
|--|-----|
| gcacccaaat cactagcact ttctggaaca tggcaggcct tctttggctt tctgctgtgt  | 60  |
| acttctgttc cttccaccct agccccaccc atcctctccc catccaagag caaacagctc  | 120 |
| tgaacagtct ggagtagctg gagacactcc tcattcttggc actctccttg ccacttgcca | 180 |
| tctagcagag ctggatgctt ccttgagcgt ctctctgtc catccccag gtatctaggc    | 240 |
| tgcctcccat ctccccact ggcatttgaa ctttaagagc ctggctcttg tgcttggaat   | 300 |
| ccaatgcaaa ggcttcccat aactagcact, ccataaacia cttttgaaca aaaattcaaa | 360 |
| ttcccagtggt ttcagttgca ccaagttcaa gactaagtat ttcaaataaa aaaaaaacia | 420 |
| aaaaaaacia aaaagggtt gggcggaacc tccatgggca tctagctggt tcccgtttg    | 480 |
| tgtggtcatt ggttatccg ctcacatttc ccacacactt tcccggccca cacagcagat   | 540 |
| gtgagagaga caatatccgc gccgagacgc agcaacacac cgccacacg              | 589 |

&lt;210&gt; 63

&lt;211&gt; 212

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 63

|   |     |
|---|-----|
| taagcccttt atagcttaat tctatatatt aaattttccc agttgcgaga aaaaacaaaa | 60  |
| caaaaaaaca aaacaaaaca aaacagcgtt gggcgcggta acaccaatg gcgccccaaa  | 120 |
| cgcgtggttc ccgtggtggt ggcacatatg tggatgatc ccggctccaa caaattccct  | 180 |

acaacaaata acgggaagaa aaggccaaaa aa 212

<210> 64  
<211> 658  
<212> DNA  
<213> Homo sapien

<400> 64  
gcgtggtcgc ggcgaggtct tttttttttt tttttttttt tttttttttg ggcgcttggg 60  
ctatgtttta tttgggcaaa gtaccttata aaacataaca ggcaaataac caaaaaaaaa 120  
catccttgac ttaaggaggt gaaaaataat ctcatgaaaa agttaccact aggataagtt 180  
agtgc aaatc cttatccata aaaatactct cttaaggggt gcagtgaagc gtcggcgtag 240  
actcgagggc ttagtagcgt gtccgcgggg gtgaaagtgg tacactccgc ctcaaatcc 300  
cacacaacca atcccagaa cgcacacgga accgcaaccc aagcacacaa gcagacgccg 360  
acacagaccc gcacccccag caagccaccc ctccgcagcc caaccaacga ccaccaccgc 420  
aaccaccacg ccagcgcacc acacgcgcca cagcacacga acaccgaaa cgaaccacga 480  
aaccagcaac caagccagca aacaccaaac caacaccacg acaggcaacg caggaagaca 540  
accaaacacc aacgacaacc cccagacaac acccaccgga cgcaccacag cccaccacca 600  
cagcgcgcca cccaccagca caccggacca cgcgcggcag cggccgcccc accaacc 658

<210> 65  
<211> 226  
<212> DNA  
<213> Homo sapien

<400> 65  
taatgacata taggcgcatg gttccctaata gcatgctcga cggcgaggt gatggatact 60  
gatgcccatg tgggtgattt cagtctccag gtcaactgag atagtgtgac ccagagctcc 120  
taccttaaat catgtggttg gtcttccac tctacatcaa aatgttgcta tctgggatag 180  
cccaagatcc ccagacaaac agagattact taccaaggac aaaggc 226

<210> 66  
<211> 430  
<212> DNA  
<213> Homo sapien

<400> 66  
ttggcattag caacctcaaa aactctggaa aaggcttcat tttctccagt ctctgggag 60  
aggagaggca ccatggaagg cagacccatc cagagaacac ctgcgacagg ctgagaagcc 120  
attgggagac acacttctga acaccaccac tggaaaatca cacatgctga aatgggagag 180  
ttccctgacc cccttgagg atatgtgaca ggagtgtggc tcatctgttc agctggagtg 240  
catactcaaa ccccttatga gacaaggagt atgcagacag aagggtcagg aactgggaag 300

35

caaaatatta actagttaat ttgatctcca agagttaagc ggttttaata ttactgacag 360  
 taatatcagc agtgggtgtg gaaccccatg atctcatgaa tcatagatag caactgctta 420  
 ctggacattg 430

<210> 67  
 <211> 813  
 <212> DNA  
 <213> Homo sapien

<400> 67  
 aaatggacgt gcagactcaa atgaccgcat aaaccagatc agggaaaaac agataagaag 60  
 ccagcatgac aataaagtga aactcaggcc aagagaagac agggagagac gaggcagcgc 120  
 atcagccggt aaatagcgag cagccgacca gaaccagcaa ttacacatcc gcgagcacga 180  
 cctagacaaa cagacataga cgcatacagg cacagaaacg agcagaaggg acgagacaga 240  
 gaaaaacaag acaacaacgt caaaaagagc aggacaaaaa agagcataat caagaggaca 300  
 acaaaggacg aaagaaacag caagcgaaaa aacaacacat gaacgagggc gcaaagaaaa 360  
 ggcacaagcg aacaaaaagc gaaccacagg gagaacgagc gaacaaacag gaggcggcg 420  
 aaaagtgaag agaacgagaa taacaccata atgacacac aacgaacaca caccacgtga 480  
 cgagagaaaa cgacaacaga aacacgaaag gcacagcaaa acgaaacacg acgcgagtga 540  
 cgaaaagcca cagacaaggg cgtatacaaa ggactacgca agcgcagtaa cccaaccaag 600  
 agaaaacaca caaacagggc gagcccgcac acatggcaca gaccaccaga acgcatgaag 660  
 acgaacaaca ccgagcagca cgaagccaca agagggaaaa gcgagggcgt gctaaatacc 720  
 aacgcggaaa agtaaaacag caggaaggaa agcagaagac aaagcagaga cataggagtg 780  
 acacagacca cgaaaagaag acaatgacag gat 813

<210> 68  
 <211> 444  
 <212> DNA  
 <213> Homo sapien

<400> 68  
 caaacaaca aaaaaaaaaa aactctggtc tccttttagga tatgttaccg tgccccacgt 60  
 gcagactaga agaaattaac tgggtgtttg gaacctttt acgtgcaaac ctttgaaaat 120  
 gtgctagaaa cccaagcatt gaagaattaa attactgtgg gtgggaaaca caggggcatt 180  
 gtgcattatt gcattattac atttggttaag gttagtaag gtttaggaaa ggcatagcct 240  
 tgggtggtat tcttgaacac attgaattcc ttttggggc tcaggtgtag gaaaggcagc 300  
 agccagaatc catataggga attgaatacc ttcaaatctg gtggtctgga ggaattctag 360  
 agatttaacc cactggtggc ctatttttaa acaacaaca aaaaaacaaa acaaaaaaaaa 420

36

caggcggggg gcggaacccc gggc

444

<210> 69  
<211> 273  
<212> DNA  
<213> Homo sapien

<400> 69  
ctgatataga tgtaattgcc aaaaatatta tagaaaactg gctccggttt tcacatagtg 60  
tggagtgaat aaacacaaat ccagattcac ttcagaaaaa aaaaaaaaaa aaaaaagggtg 120  
gggcggtaac catggccgac agctgggtccg tgtgtgaaat gggttcccggt ctcccatccc 180  
catttcgacg cccaaaaagg aaaggggaag aaggaagacg gacaacgaag ggtcagaaag 240  
gaggcaccag cggcagaggg aaaagctacg gga 273

<210> 70  
<211> 1397  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (255)..(255)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (259)..(259)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (325)..(325)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (354)..(354)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (356)..(356)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (623)..(623)  
<223> a, c, g or t

<220>  
<221> misc\_feature

37

&lt;222&gt; (628)..(628)

&lt;223&gt; a, c, g or t

&lt;400&gt; 70

```

gcgtaggtcgg gccgaggtac actcttcccc tctcgggttcc cacaggcaac gttaccatca      60
gaaaaaata agtttcaggg gccaggtattg gagggggggg ggagcgaggg gatattgtggg      120
taaaaaccag gtccaaatct caccaataga ggaatttttc aaaatagagg ttattccac      180
attagatcca tctcatcctt cctctccctc tatccttcag aggttcctct cgttttcgcc      240
ttctctgtaa cccncttnt ctcttcttcc taaccacaag cctctcttcc ttctaattctc      300
ttctctctgc gtctaattct atacctctc tctccaatct gggtatata accnctat      360
ctcttcta atcccatctc ctctcaactct cactctctct cacacactct cacaggctctc      420
gctctcgtc tcttctcaca ccttctcac tctcaactct actctcaatc tcaactctggg      480
ctcaactctc tctgggtctt tctcccat tatcacgtgt gagacacatc tcttcccatc      540
tcataactc tgcgtctcgc tctcaatctc gctctccatc tccctctctc ctgcgtctca      600
tctcatctca ccagaggggc cctctctncc acaggtatag acgccccctc tcagacaatt      660
ctccggagag tctcaggagg gggcgccctc tcaactgtgt tctcgggtct ccccgggcg      720
tctcaatatg gcgcgggtct cggagacgat cacttgtgtg tgaagagttt gccgcgggtg      780
gagagggaga cctttgtgac acccacacca attttttct ctctgggggg ttagagttct      840
cgagtctccc agaagggttt ggggggttaa aaacctctg cgcgcaaat ctgtgacaca      900
caagcgggtt ctctataaga gcctccctt gggacgaggg gttctatttc cctaaaaacc      960
tttttttcc acgagggggg gccatcccta ttttggggg gtgccctgtg aaggggggtc      1020
ctctttaaac atcttctctg tgttttgggc ccacctttt ataaacattt ttaacgcaca      1080
tgtgcccttg taaaagggtt ttcggcgaca ccacctttt tattactcag ggccccaat      1140
ttataacctt tccccagag gtgccccccc cctctctga agggaaaaac ttccctgagg      1200
ttaattaccg gggtatttaa gaggtttcaa aacaggggcc tttggagggc ggggttaaaa      1260
ttccaattgt ggggctcgcc aattaaaggc ctgggggtgt tccccctggg gttggttggc      1320
gacaaaacat tcgggggtct aatccccggg gctctcacca aattcccccc attcctcaag      1380
cgacccagac ctacacg      1397

```

&lt;210&gt; 71

&lt;211&gt; 844

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (595)..(595)

&lt;223&gt; a, c, g or t

<220>  
 <221> misc\_feature  
 <222> (644)..(644)  
 <223> a, c, g or t

<220>  
 <221> misc\_feature  
 <222> (695)..(695)  
 <223> a, c, g or t

<220>  
 <221> misc\_feature  
 <222> (758)..(758)  
 <223> a, c, g or t

<220>  
 <221> misc\_feature  
 <222> (783)..(783)  
 <223> a, c, g or t

<400> 71  
 gcggccgccc tgggcaggtc ccccccttt tttttttttt tttttttttt tttttttttt 60  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 120  
 taaaaaaaa aacacccgag gatgatgagt gggggggggg agaagaaga gaaagaaaga 180  
 atagtggttg gtggttggtt gttatataaa aaagatggtt gtggtggttg attgtgatgg 240  
 agagaagagg aggtgtggtt cttgttggtg agatagtgtt ggggtgtggtg tggagggtcga 300  
 cacaccagc acaggcaggg tggagtgcg tgaatcagct atctgagaga gagagagagg 360  
 agagagtata tatgtagggt gtgtgcgtga cacacaaatt ataatgtgta gtgtgtgtcg 420  
 tctcgtctct gctgctgaga gatgagagag agagagagtg tatatatatt gttgatacac 480  
 acacacacac acgacacat gcgtcgtgtc gtagtcatca tcaacaacat caacaacaaa 540  
 ataatagtaa tagtagtcat cgtctcgag cagcagcgcg agaatgatga tgatnagagg 600  
 gtagtagtgg tggtagtgat gatgatgaga gtgatggtat acgntctgta gtgtcatcag 660  
 tgataggtag gtgtagtga tcatcatgat gaganaagaa cttaaataata atgatcatgc 720  
 atcatcataa taattattac tagtagttcg tggtaggtng tagggaagat ggtgcggagc 780  
 aanatagaga agtaagagca gcaggtagct gctgctgctg ctgactgatg actgatgatg 840  
 atta 844

<210> 72  
 <211> 738  
 <212> DNA  
 <213> Homo sapien

39

<220>  
<221> misc\_feature  
<222> (327)..(327)  
<223> a, c, g or t

<400> 72  
gcgccgccc gggcaggtgg acgttggtgt tagaggttag catacgcac aaggcacaag 60  
taagctacca tggactcccg caagttttgc caataaccct cgtgatgcgt ggccttctgg 120  
tcacagcgcg tctcgcagaa gatcactatg gctgtagcat ttcagtcgct aatcccgtgg 180  
gggttgacgc tctgtgtcaa taaagttgcc gctgatgagc ttgtactcac aaggaaaatg 240  
aaggctaagt acgcaagtat ttctagcaga caacatactg attgatacga atgacatacg 300  
attatagagt ggacgatgaa cgagaanggc taggatattct ttgtcaggaa gtagtcaatg 360  
tcattcggtg tgaataatca caagaatttt ctatacgagg ttggattata ccataggaag 420  
ttatggctga cttgacttgt gtggtatcct tggaacatca tagactacaa tagaatcatg 480  
tagggctaaa aggaaagact aagctttccc ttcctttgga agtaaacatt aaaaaccaa 540  
ttataacaaa aaaccgaaaa gagaaacaac atacaacaga acatcaacaa aacagagacg 600  
cttggggggg aaaactctcc gtggggctca atataggcgt tgtattcccc cgcgtgtgtg 660  
gtggaaaaat gtgtggttat actcgcgggg ccacacaaaa ttctcccaca cccaaatttt 720  
tcggccgcac gcaaaagg 738

<210> 73  
<211> 292  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (236)..(236)  
<223> a, c, g or t

<400> 73  
gactaagcat aatggcgact ggcccatcta atgctttgaa cggcgcagtg tatgatgaag 60  
ctgaggtgga ggatacttaa gccaggagca gaggtcacia tgaagcgaaa tgtgcaactg 120  
cactccagcc tgggcacaga ggaagatctt cacagaaaaa aaaaaaaaaa aaaaaaaagt 180  
ttggtacatg gcatctgtcc ctgtgtgaat gtatcgcggc aatcccaata agaagncgcc 240  
acagaataga gagaaataag ggaacaataa taccaagcga agaaaggaaa ta 292

<210> 74  
<211> 785  
<212> DNA  
<213> Homo sapien

<400> 74

40

```

agatcatata gggcgactgg gcctcctaata catgctcgag cggcgcgatt gtgatggata      60
ggcggcgccc gggcaggtag ataaggtaaa aataaaatcc taagccccc attgacaaaa      120
gggaccttct cctgaccaag gggatcacca gaaaaacctc aacctgaat tcccagaaca      180
tgatgggatg ggaggatcatg atgcgcctgg taatagcccc ctgtttcaga gatttggtac      240
taccacaatc tggggcggcg attcatgtta aaacagagat cgtaagactg acagaacgga      300
ctctgtggca ataagatacc aaattataaa caggacccaa agccatgcta ggcgagggta      360
agtcaggcaa cccacactta gagaataaac tatattctaa gagccacaag gctttctgtt      420
tctctattag ccaaacacac actagccttg ggatagggaa tattaaaaca attgcagctc      480
cactaggtag caactaactg actctgtttc accagccata gcagctgtga ttggacaaga      540
gactgatttc agtgactttc tccataataag agaccaccga cagctgacat gccgacagct      600
gacccgttaa tagagagaga tgatgcacct gcatgccttt gtgtctgaaa agacgtttgg      660
cataaaggcc ctaattgtag atgtgtaatg taagtctcca cccaagtga catgggtcct      720
atcttcatgt tgctcaaaaa ggggtgtgtcg ggcacttatg aatatagtc cggtacgtga      780
ttgtg                                           785

```

&lt;210&gt; 75

&lt;211&gt; 1226

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 75

```

ggcttctttt ttcatatgac atgtatctac catcctttga gtacttactt attttctggg      60
acaaccagat gttcaaggat cctccccttc tctgccagg cctggcatca gccattgttg      120
gcaggagata atttgagcag atcgtgtgga ttccagaagc atgaaaaacta ctgtgaggat      180
taaataagtt agcatgtata acattctggt gcttttggg agtttccaaa ttgtcatgaa      240
caagcactac ttatagaca ggaaaaaaag tgattcaaaa tgtgaaaacg ggtatagtga      300
aaaataaaat cctaagcccc ccattgacca aaggacctt ctctgacca aggggatcac      360
cagaaaaacc tcaacactga attcccagaa catgatggga tgggaggta tgatgcgctt      420
ggtaatagcc ccctgtttca gagatttggg actaccacaa tctggggcgg cgattcatgt      480
taaaacagag atcgtaagac tgacagaacg gactctgtgg caataagata ccaaattata      540
aacaggaccc aaagccatgc taggcgaggg taagtcaggc aaccacact tagagaataa      600
actatattct aagagccaca aggctttctg tttctctatt agccaaacac acactagcct      660
tgggataggg aatattaaaa caattgcagc tccactaggt gccaaactaac tgactctgtt      720
tcaccagcca tagcagctgt gattggacaa gagactgatt tcagtgaact tctcctgata      780
agagaccacc gaccagctga ccatgccgac cagctgaccc gttaatagag agagatgatg      840

```



41

```

cacctgcatg cctttgtgtc ctgaaaagac gttttgccat aaaggcccta attgtaagat   900
gtgtaaatgt taagtctcca ccccaaagtg aacatgggtc atatattaca tgctttgttc   960
aataagaggg catgtgtcag gaccaccttc atgaatatc atagctctc ctgttacctg  1020
ttgaatatgt atgttttagc aatcccttca gcatagcgct cctgccccaa cccctectcc  1080
ttggacgtgc ctgtctctgg ccttggtgag agacagattc ccagcctcag acagatggcc  1140
gccaccttgc aggctacgac cgttttacaag aaataaagcc ttctcttttt ccaaaaaaaaa 1200
aaaaaaaaaa aaaaaaaagg gcggcc                                     1226

```

```

<210> 76
<211> 792
<212> DNA
<213> Homo sapien

```

```

<400> 76
gcggcgcccc gggcagggtt tttttttttt tttttttttt tttaaaaatg gagtctcgct   60
ctgttccccg ggttgaattg caggggtttc atttgggttc acgtgcaacc tccacccccg   120
ccggttatca agaaattctc tgtgcctcag ccactcctga aatagcgtgg gaccatacag   180
gacccccata accacgcccc agataattga ggcgtattta taataaaaaa caagggtttc   240
acacacatgt tatggcccag gttgtggttc tcaaatctct gtgacctctc aggtgtgacc   300
tccaccgtgc ctteagagct ctccacaaca aggtgcggggg attacacggg gtggaaggc   360
caccacaccg cggccttgac aaattgactt gtggagctca tcagtttagc gcactcaaaa   420
agttcaacaa atttaggcga acatttctca aaattacaag agattatagg cgctacagga   480
gaattgtaca cacattttca atatagtgtc cacagtggcc gtagttctgc atgtgggggg   540
aaaaaatata gggcgctcaa ttaattagat gttcaccatt caccgagtga ggatccccc   600
taaaattttt aggegaccac atatacttat tgggtcgtg ccaattcctt cattattccg   660
agggcccaaa cttttcttta ccagctcctc agcgatcatg ggaaaccctt ttgtagttta   720
caccacaag aggggttgga ggtggaataa gcccctttac gttatgttgc ttatgaaggt   780
gatatcgcta tg                                                     792

```

```

<210> 77
<211> 946
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (177)..(198)
<223> a, c, g or t

```

```

<400> 77
ttgcaattgc attggtgctt gtggatggcc atctctgttg atttttgtga tttgggttgc   60

```

```

ttgtgtttta ttgaaagga caaatgagag aagtgccttt catataattt tatacctttg 120
caaatgggtt aaacttttca tttgatcaa gaagatgcca ttgtttaaaa tggtagnnnn 180
nnnnnnnnnn nnnnnnnnga aatggagtct cgctctgtcg cctaggttga attgcagggg 240
tttcatttgg gctcacgtgc aacctccacc ccgcccgtt atcaagaaat tctctgtgcc 300
tcagccactc ctgaaatagc gtgggaccat acaggacccc cataaccacg cccagataa 360
ttgaggcgta ttataataa aaaacaaggg ttccacacac atgttatggc ccaggttgtg 420
gttctcaaat ctctgtgacc tctcaggtgt gatctccacc gtgccttcga gcttctccac 480
aacaaggtgc ggggattaca cggggtggta aggccaccac accgcggcct tgacaaattg 540
acttgtggag ctcatcagtt tagcgactc aaaaagtca acaaatntag gcgaacattt 600
ctcaaaatta caagagatta taggcgtac aggagaattg tacacacatt ttcaatatag 660
tgtccacagt ggccgtagtt ctgcatgtgg ggggaaaaaa tacagggcgc tcaattaatt 720
agatgttcac cattcaccga gtgaggatcc ccataaaaat ttttaggcga ccacatatac 780
ttattggctc cgtgccaat ccttcattat tccgagggcc caaacttttc ttaccagct 840
catcagcgat catgggaaac cctttttag ttacaccca caagagggtt ggcaggtgga 900
ataagcccct ttacgttatg ttgcttatga aggtgatatc gctatg 946

```

```

<210> 78
<211> 895
<212> DNA
<213> Homo sapien

```

```

<400> 78
tgggtcctct taatgcatgc tcgagcgtgc gccagtgtga tggatgcgtg gtcgcggccg 60
aggtcctcc cctttttttt tttttttttt tttttttttt tttttttttt 120
taaaaaaaaa ccccgattt ttgggggggg ggggggaaaa aaaaaaaaaa ggggaatgtt 180
tttaaaaaaa agaggggttt tctccactca gggtgattaa aatgaaggag tatatatgtt 240
gtgttgaggt ggtgggtggg ggtgaggtgc accccacatg tgggtgctgt gggacaaact 300
attgttaaga agtggttaata ttagggcgtg ctacactact ttacttgttg cacctccgca 360
aagaagcagc ataagtattt cttttgtgta acacgaaaac aaactgtgtg gctccatcca 420
cacaccacac ataataattt tccctcccca gtagtgatta aaaataagtg gggggggtaa 480
ataggcaaca gtttttcaac gcaaaagctg ttgctccaaa aaaaagtttc tccacaaaaa 540
tagtcttctt tgagtggggc ataactaata tcgttggaac ctctcctgt agagaagaag 600
atatatttat attacgcgca cagagtgtgt gaaatcgagc gcgtctttcg aagaagtatg 660
agtgaagtgt tgactgcacg gcgggaagac aaatataatt ctaatgtgga cagaattatt 720
aatcctccgg gcgggcgcca ctattattat aaaaaaatat tcatgtcggc ccctgtaaaa 780

```

43

actacttgtg gggcataacc acaatggggc aaaataaggt ttttccctg ttggtataaa 840  
aattgggtta cctccgcgcc caaatttcca caatattgtc gacacacaac aacct 895

<210> 79  
<211> 1049  
<212> DNA  
<213> Homo sapien

<400> 79  
gcagcacaga aaaccagcaa aaacgcagtg aatatcacta tagggccctg gttatctata 60  
tcatgtctga gcgcgcgcc gatgtgatgg atgccgccg ggcaggtcag ctacttggga 120  
ggctgaggca ggataatcgc ttgaacttgg gaggcagagg ttgcaatgag ccaagatcgc 180  
gccactgcac tgcagacctg ggtgacagag caagactcca tctcaaaca acaaacacaa 240  
cagggcataa ttacaagccc aacgtgcgtg ctctgaagga aggcgacctg tcagcaactt 300  
aatatcccaa ggatctggcc ggggtgtgtg ctggcatcac agcctgttaa tctgacgcc 360  
ttatggcggg gccaaaggtt ggaaggatca cttgacgcct cagagagttt cagcgaccaa 420  
gcgcgtggcg gccagcaa atagataaggac cctctcattt tctacgtgtt gtcatacaca 480  
tctcactaaa aacaaacaac aacaaaacaa ccaacaaacc aacgcccattg tgtacgacgg 540  
taacacgtag tgtggcgcat acccatcgtg ctttccccag gtatgacacg tgagtcacca 600  
cgaggacaaa agtgggccgac ccaacaaaat ggcgcagaag aagacgccga ggaggagaag 660  
gagggcgccag acggcgacac acaaccgacg cggtagcacg acacgagaag acgacgaggg 720  
caggagccgg aggagaggaa ggcgcacgac aggcagagcc atgagcgaga atggaccaca 780  
ctaagcaca gcaacggacg agtcgcccga gcggaggcaa caaagagaag cgacagacag 840  
cgagggctag agcagagcga gacagagaca gccatagacg cagcaaaaca acgagcagaa 900  
agagcagaga aagatcaaag gacagcaggg acgcacagag acgccagacg cagcacagac 960  
ggccgagcgg agagtgtcac agcggagcag gcggaagaca gcaggccaag agaggaacag 1020  
tagcggaggg actcctaata gaccacgag 1049

<210> 80  
<211> 840  
<212> DNA  
<213> Homo sapien

<400> 80  
gcgtggtcgc ggccgaggta cacattaaga atgtgcaatt gggctgggca tgggtggctca 60  
cgctgtaat cccggcactt tgggaggccg agacgggtgg atcacaaggt caggagatca 120  
agaccatcct ggctaacacg gtgaaacccc atctctacga aaaatacaaa aaaaaaaaaa 180  
atthagccag gcttggctcg gtgggcacct tgtagttcc cagcttactt caggaggctt 240

44

```

gaggcaggag aattggcggt gaaccttggg ttgatggagc ttgcagtgag ctgagatgtg 300
tggccacgtg cactccagcc cgtgggctaa cagagttgag actcgtgtcc caaaaaaaga 360
aaaaaaaaa acaagattcg tgccaatgga gtgtgttttc tgaaatttta tcctgaagct 420
tggtgaaaaa tttttcaaac aaatgtgccg tgaggttttc ccaccagggg ttgtgacact 480
tattttaaaa ttccctgtgt cagccactgg tttgttgaag aaattcctac gtggctctac 540
cacattcttt cacccaaaca ttggcatcta caactaaagg tgccctttta aatttaaccc 600
attttgggtt gcgatcggtt ggtagtgggt gtccggccat tggggcggtt tatccacct 660
tcggacatta accggaatgg cctaagggat tattaagcgt cccctttttc ctttttgacg 720
acacacactc atacacacag cgaaaacggc ttggggcgac acccagggcg ccaaaacggt 780
agtctccggg tgtaaatgg gtaccgggc caacaatccc caacattact cagcacacag 840

```

```

<210> 81
<211> 864
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (568)..(568)
<223> a, c, g or t

```

```

<400> 81
gcggcgcgcg ggcaggtccc ccccccttt ttttttttt ttttttttt tgggagaggt 60
aaaaattttc ttttattcca cggaacaaat gttttattat ttaaaaaagg gggtttttt 120
tttttaacaa tttttggcga aaatttatat cggagatagg ggggttaaac ccctgggata 180
gcgctttggg tataatagtt cattatcagg gggcagatat tattaggagg aacaaagggt 240
acaaatactg gagtttgggt ataaacatc ataatttat ggggtcttgg tgggagatta 300
taaaccgcat tacaccctc tcgtgttaca caccggtgga ggcaattaaa ttgtgtggca 360
gctttccacc aacacactaa agtgggtgtg gctttctcag taacacacgt ggttggagga 420
acatccacat tctttttcgt gcaagaaggc ccctgcagtt tctacaaatt catgcacccc 480
caaaccatct cctccttatt tctctgtgct atacatttat ttataaagcc atatttatat 540
attttttctc atacgcccaa ctgcgggnct atagaataaa ctccataagt gggcataagc 600
attattcggg ttccgagtgg gttattctc aggtgtgtaa tatctataga tatgtggtgg 660
ggcggcggtg gcgtaacact acgggttaagt caccaaattc gttttatata gttaccccca 720
aaatggggtg gtggcggtta aaacttctg gcagggtatt aagactgtgg tcgcttaaac 780
atctatcggg gctttctcta caaagggacc ttaatacgt tttattgtaa tccctggagg 840
gttgaaggga ccacataagg tatg 864

```

<210> 82  
 <211> 896  
 <212> DNA  
 <213> Homo sapien

<400> 82  
 gcggccgacc gggcaggtgc cagcgcaggg gcttctgctg agggggcagg cggagcttga 60  
 ggaaaccgca gataagtttt tattctcttt gaaagataga gattaataca actaccttaa 120  
 aaaactacta gtcactacgg ttacctacac gactacttgc ttacggcggt aagtttttta 180  
 tagcgtagag ttgttacata cgccttaacg acttcttaac gagacgaact actgacggga 240  
 ccttacgaca cgacgctagc cctgacgcga acggacaaca cgactagcaa cggttctctt 300  
 caaccaccag ttgcacgtga cgggtctgca cgactgcaag cgttcgcgcc gggtcagcgt 360  
 cactgcgcgt ctactaacgc tcgctctctc gctcgcgtgc tcgcaccgac tccgctctca 420  
 ctccctggct tccagcggcg gtgtcgccac agccacctcg tactcgccgt atgtcgatgt 480  
 cctgtggtgc gggcgcgcc ctccgggttt gcgtgctggt gtggctgtgg gtgggggggc 540  
 gtgtgggggc ggtggtgctg ccgcgtgcgc tgtggtcggc gtggggggcg gtgggtggcg 600  
 gcttgcctctg cgtggttctt ctcttctggt tgtgtgcggg gcggcggggg gcgcggctgc 660  
 cgccgtcccc ctgcggtgct gttgcggttg cggcgggtcga cgccggcgcg gcggggggcg 720  
 tgggtgcgtgg tgggtgggtc gtcgtggtcg ggcgttggtc tgggcgcctg ggggtgggtg 780  
 tggggcgggg gtgtgcgcgt ggtccttctc tgtgtcgcgg cgggtgcgtgg gcgggcgcgg 840  
 cggggcgcgg gggggcgggg cggcgggggc gtcgcggccg ggcgcggtgt cctggg 896

<210> 83  
 <211> 954  
 <212> DNA  
 <213> Homo sapien

<400> 83  
 ctagatccat tgtcgagcgg cgcagtggtg atggatgtcg cggcgaggtc ctcccccttt 60  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 120  
 tttttttttt tttttttttt tttttaaaaa aaaaaaaaaa ctctcttcta tcaaaaaggg 180  
 gggggggggg gcccgcgggg gggggggggc ggcccccggg gggggggggg ggggggaaaa 240  
 aaccacaaaa aagaggcgaa caacaagcgg gccgtagtca caccacacca ccccaggcg 300  
 caccaccccc caccctggg agaaagagag ccctctccga gagagggaagt cgtcgacgca 360  
 ccctcaccaa acgcgcccc cccccaaca ataatacaca aatacgcgag acaacaacca 420  
 cgcgccaccc acaccgcggg cgggcgcgcg tccctctctc tccgtcttcc cttctccgc 480  
 gccgtccaac atcacactcg tgcctctcat ctctgtactc ctctctgagt gcaaagacga 540  
 ccacccgacc cccctctacc tccccccca acaccgaagc gtgcggcggt caatccaccc 600

tgctcaacaa aacactatcc ttccgccctg cgagcgcaga attccttctt cgccgccgat 660  
caacatcccc cacaatatata actcctacga cactctcatc cctccccctt tctcctccct 720  
cacctccatc ccactcctcc ccccatcccc cctccactcc actccatcct ctactccctc 780  
cctttccctt acacctctcc cccactcac ctatctctcc ccacctaca ctaatccata 840  
cttatcaciaa ctcatctca ctcttcaatc tcaactaacc tcaactctac ctctccacc 900  
atactctcta cccccaccc caccacacac caccacactc ccacctcaat acac 954

<210> 84  
<211> 918  
<212> DNA  
<213> Homo sapien

<400> 84  
gtaagagagg aaataatata tatagggcac tggttcatct agatgcatgc tgcagcggcg 60  
cagtgtgatg gatgagcggc gcccgggcag gttttttttt tttttttttt tttttttttt 120  
tttttttttt ttttttttgg aaaaaaaaaa ttttaaaaac ccccaaaaat ttcccgggca 180  
aggggggctt cccccggga aaaaaaaaaa aaaaaaaaaa atttggggcc tctgggggtt 240  
acccctctcc ctagtggggg gataaaaaat aaccacacaa taatcacctc ctacgatca 300  
accggccgcg ggaagacacc aaagcagcgg gggggggggg ccacccaca gctgaaccg 360  
gtgggggtggc aggggagggc ctcggtcggt ggagccccgc gtggggagac agcagcggaa 420  
aacaccccc caaccacagc ggtggacgag aaaaccccc ccagagacg ggggagcgat 480  
ctccctctt ctccctatag aacgcctcct ctctaacaca cgcgcggagg gccccggt 540  
aagctcccaa agaaaaatct atctctgata gagagtgaac acccctcgat ctacttcaa 600  
gaaaagagtg aagaagagac ccgcgcgcac ccgagagcaa cgcgagagtg aagcgcgaa 660  
agacgaacaa gagaacgctc gccgtgcgtg agacacgtag agaaccgccg ggtggaggag 720  
aagaggagag atatcatacc tccctctctg gtggggaggt atgtgggcgc gcgtccaga 780  
ctgcttgctg gcgcgagacg tctcatgtga gcgacaaaa gccagtgtgc acccctgct 840  
gtgtgaaaga aatttgctg ttcctcccc cgacacaaa attcctccca aaaattataa 900  
ctgaacaaaa ccaaccgg 918

<210> 85  
<211> 728  
<212> DNA  
<213> Homo sapien

<400> 85  
gaggatgac actcatatag ggccatggtt ccatctagat gcatgctcga gcggcgcagt 60  
gtgatggata gcggccgccc gggcaggtct tttttttttt tttttttttt ttttttaaaa 120

47

```

ggggcaaaaa ttctttttat ttattccatt ctcccccaaa attagcataa taaaacccaa 180
gggaggagga ggggggtaga aggtagacaa gatagagtct gggaggaccg aaaaagggtg 240
gtagtgcgcc ccgtggaaaa ggttgtagag aggccaaatg gatggggagt ggtggtacag 300
tgcttgaccc tagaatgagc acgtgggggc acttctcccc ttctaacatc ttctcccctg 360
ttagaagtct tctttgtaga aggggcgatg ctcaaggccc tggaatgggg tgagacattc 420
agaaggctgt aaaactttgg tgggtctcatc gaaatctggc ttcgagcacc accgtaaggg 480
gtgcggcaaa gaggtggaag tgtctcggcc ctggagtagt cctggcttct gtgacactct 540
cctgggagag gtacaccggg atgggggggg ggcgtaacac aacggctggg gtggacacca 600
tggggcgcat aagactgggc cccggtgtgg ggagaatggg ttaccccggc tcacaatccc 660
ccaaaaataa tggcgaaaca atcagacaaa actcccgctg agacagggaa cacaagacaa 720
cataataa 728

```

```

<210> 86
<211> 265
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (198)..(198)
<223> a, c, g or t

```

```

<400> 86
cttaggaaaa tcaaggccgc aaaggcaaat aaatcttgtt tgtcttcacc catgtgaaaa 60
aaaaaaaaaa aaaaaaaaaa aaaagttggg gggtattcta ggggtctata agtctgtgtt 120
tctgtgtgtt gtggaatatt gtgtttatcc cggtctccac atattccaca cacacaatct 180
attacggaag cacaagcncg acagacaatc aacaccgatc acgtcgtata tctataacca 240
gagacgtagg cgacacacga ctcac 265

```

```

<210> 87
<211> 430
<212> DNA
<213> Homo sapien

```

```

<400> 87
tgggccacta gatgcatgct cgagcggcgc gggcaggtcc cccccccctt tttttttttt 60
tttttttttt tttttttttt tttttttttt tttttttaa aattatatat tttaaaaagg 120
ggaatttata aaaaaaaaaa aaaaaagggt tgaccccaaa aaaaaattaa aaagtggggg 180
gcataatctc gggggcaaa ggtgtaccgg tggggacagg tgttaccgc caaaaattc 240
caaaaaacaa acgaagaaaa aacagacaga gcaaaagaag cagagcacac cactgcagcg 300
cgcaccagc aaagatagaa agcagacaag agacatatcc ggtgccggaa tcacctggg 360

```

cagacgcggg gtcggcggcc atccacgcgc ctcaccacac cacggcaaca acgcgccggc 420  
gggcgagggg 430

<210> 88  
<211> 868  
<212> DNA  
<213> Homo sapien

<400> 88  
gagcggccgc ccgggcaggt ggcagcactt gtaaaaataa agcagtaagc aaaatccttt 60  
taaaaaaaaa aaaaaaaaa aactcggaaa gaaaaaaaa aaagaaaaaa aaaaaaaaa 120  
aaaaaaaaaa taaaaaaaa agaaaaactg gcgcacgatg tcagggcaca tctacagagt 180  
gccaggggaa cgtgggccac aagattcatc aatgggggag catccagtcc agatgacaga 240  
ccacagttaa acaagcatca cggaaactct tatgacatac atcatggata aactagattc 300  
cagtaggtat ggaaccaact gggtgaaacc acatgtccaa acatactagc aagtaggcac 360  
agcaacaggc ctatgaatag tgatccgccc ataacagtgg gcaagcagcg actagaaaca 420  
cactcctcaa gcaaagtcca agcagcaaga gaaagagcca tcgaatagga gacaccgggg 480  
aagaaaagaa caccatagct aaacaaacat acagacaggg aaaagacaag cgttaaacga 540  
tgtgagaaag gaaaagaata tagaagtata gtcagtcgaa tatatatata agctgcacga 600  
aaaatttaga acataataaa caaacaagag agatgtcaca tatatggggc agccaaatat 660  
atttcagaga tgttgccata aatgaagttc aacatacatt taattgcaga gatgttacct 720  
ataaaatggt gtaataaaaa gagataataa ggaatgaata ctttaaaaaa gatatatattg 780  
ggctagaaga ggaagacaca aaaaaaaaa cagaaaaagg gaaaatatag cgggaagagc 840  
agaacagagt gaaaaaggaa aaaggtag 868

<210> 89  
<211> 1682  
<212> DNA  
<213> Homo sapien

<400> 89  
ccacggaagc ccttttcacc taccctaaag gagctgdaga gatgttagaa gatggctctg 60  
agagattcct ctgcgaatct gtttttagct atcaagtggc atccacgctt aaacakgtga 120  
aacatgatca gcaagttgct cggatggaaa aactagctgg tttggtagaa gagctggagg 180  
ctgacgagtg gcggtttaag cccatcgagc agctgctggg attcaccctc tcttcaggtt 240  
gatactgcct ggatggtcac ctctggtgcg cagcaagtgc aaagccagtg ggggactttc 300  
tcacagctta catagccatc cagagatcca cagctacgtc actgaattgt taatgcacat 360  
ttgtacttgg tttctctgta tctattcaca ggcaacaaat acttatatgt gtgatctttc 420



49

```

agggaatgtt ttgtttatTT gtttttaaaa gtattgggaa tcagattaag acaatcagtt 480
tcagagaacc aggaggTTtg gggTTaagag ataactcaaaa attttcacaa gccaaagtagg 540
gcatatatca gatttggcca actgaatggc gtctgtcctg tcatccatat ggtgcctgga 600
aatatttacc agtcaaggTc aaggTcagca tctgtggTTa aaaatatagc attctgacct 660
aaaaaagtTa ttttgcagat gaatgtgttt tcaactcagg acctatccaa atgaggaatt 720
tttaaatatt cttttttttt tcctattttt agacatcaat tctatagatt ctgacttttt 780
ctaacctctt atagacatgc caaatgctgg caaaaagaag tgctttttgg atatggcagc 840
acttgtaaaa ataaagcagt aagcaaaatc ctttttaaaca cagaaaaaaa aaaaaactcg 900
gaaagaaaaa aaaaaaagaa aaaaaaaaaa aaaaaaaaaa aaataaaaaa aaaaagaaaa 960
actggcgcac gatgtcaggg cacatctaca gagtgccagg ggaacgtggT ccacaagatt 1020
catcaatggg ggagcatcca gtccagatga cagaccacag ttaacaagc atcacggaaa 1080
ctcttatgac atacatcatg gataaactag attccagtag gtatggaacc aactgggtga 1140
aaccacatgt ccaaacatac tagcaagtag gcacagcaac aggcctatga atagtgatcc 1200
gcccataaca gtgggcaagc agcgactaga aacacactcc tcaagcaaag tccaagcagc 1260
aagagaaaga gccatcgaat aggagacacc ggggaagaaa agaaccacat agctaaacaa 1320
acatacagac agggaaaaga caagcgTTaa acgatgtgag aaaggaaaag aatatagaag 1380
tatagtcagt cgaatatata tataagctgc acgaaaaatt tagaacataa taaacaaaca 1440
agagagatgt cacatatatg gggcagccaa atatatTTca gagatgtTgc cataaatgaa 1500
gttcaacata catttaattg cagagatgtt acccataaaa tggTgtaata aaaagagata 1560
ataaggaatg aatactTTaa aaaagatata tttgggctag aagaggaaga caaaaaaaaa 1620
aaaacagaaa aagggaat atagcgggaa gagcagaaca gagtgaaaaa ggaaaaaggt 1680
ag 1682

```

<210> 90  
 <211> 959  
 <212> DNA  
 <213> Homo sapien

```

<400> 90
ttgggttatc taatgcatgc tcgagcggcg ccagtgtgat ggatcgagcg gccgcccggg 60
caggTctccc ccccttttta tttttgttat ttggTTTTta ttttttttTc tttgtgtttt 120
atatttgTtt tgtttgttta tatatttctt attattaatc ttgtTgtTgc atatatttct 180
tttgaatta atttcattat cattgtttgt ggcattTtga tctattggta gcctatggag 240
ccatgagcca atgaggatat atagagaaca agagctgcat gatataaaa aagcctggca 300
agcagcaatc atcagacaca caacaggagg aaggTgtata ttcccaggga gggagtggTc 360

```

50

```

agtccecaag gacccagtca gctgccatca gatctctgga ttctgaaaac ataactggca 420
tcaacactgg ggtgtaagaa acatgctatg cactataatt gtatcagagg acatagctac 480
agcagatccc aacgagataa tcattccggg aaactatata cttctagcaa caacggcaca 540
ataaggggat catttcatta catatttccg agtctctccc tcggcggcta gcgagacaac 600
atcataggca cgacaagctc ctatgactgt tactttgccc aggcattgagc actatgatga 660
catgcgacaa aattcaccac gtctccatat cgcaatctct acaaatataa tcacacaacg 720
agcccttaat gcaacagtcc catcccact ctttgataag cctcgggaac ataacagctt 780
acaccatgaa caacccttg cgctacgagc attcttcaca tcactcgggt gaaaacagca 840
tccttctaac tgtaaggccc accgtcttgt tccctagggc atctgtcgag ctccagaatc 900
ggccctcctg cgatcaacct tctcaacggc tcattgtcca atttgtagcc cttgattcc 959

```

<210> 91  
 <211> 737  
 <212> DNA  
 <213> Homo sapien

```

<400> 91
gagtgatcac tatagggcgc ctgggtcctc tagatgctgc tcgagcggcg ccattgtgat 60
ggatgtctat agtgaactg tttgagacat atcagatgga gaggaatgct atgggaacaa 120
gtcctaagga accaggaaga cactggggat caagatacca gggaaaagt agcttttaga 180
gaagatggca tttctttctc tgaggataga gggctaggca cgtagagaca cactttgagt 240
aatataagtc ctttgttggg aggaagcaat aaggattggt agagaaaatg tggagaatth 300
tctgagcaat gattttcact ttattgcaat aggcccttct atcgaaagaa taaaaaatgg 360
aatttcaaaa actgatcaaa gcaaaatagc caaactgaag caggaggaaa gctagagact 420
cacacatgag ggtggccccc acattgctgg tctaacatcc aggcacataa accactagta 480
aaaggcacac aaagactgaa taaaggcttt ctagaaatgg gtagtgacag cagcatcctc 540
cattctatth cttcacttca gaaatagaag tcaaaaacac tgattttaag tgattcataa 600
ttgaaaaaca atgtcataca ttcaaggagc cttgagattt tagattaata ccataaagga 660
aaactggaag ggtggaacag ttagaaatat cacatcacat ctagaagtgc aatgagacta 720
gactgcatag gtgatgg 737

```

<210> 92  
 <211> 601  
 <212> DNA  
 <213> Homo sapien

```

<400> 92
tgcgcaaccg tgaatgatca ctatagggca catgggttat ctaatgcatg ctcgagcggc 60
cgcagttgtg atggataagc tggggcaggc agatcatgtg aggttgggag tttgaggtca 120

```

51

gcctgaccaa catggtgaaa acctgtctct actaacaata caaaattagc tgggtgtggt 180  
 ggtgcctgcc tgtaatccca gctacatggg agtctgaggc agaagaatcg cttgaacccg 240  
 ggaggcgggg gttgtggtga gccgagattg cgccactgca ccccgagcctg caacaacagt 300  
 gaaactctgt ttcaaaaaaa aataataatc aaaaaactta gccagacgtg ctggcgacaca 360  
 cctgtggtcc catctactca ggaggctgag gtgggaggat cacttgaaac tgggagttca 420  
 agtttgcagt gagctatgat cccccacta cactccagcc tgggcaagag tgacacccag 480  
 cctaaaaaaa acaacaaaaa aaaaaaaaaa aaaaacacct gggggatacc ctggggcaaa 540  
 ggggtgttccg ggggtgtgaca aatggtttcc ggtcaaaatt cccccaaaat cgcagaaaag 600  
 9 601

<210> 93  
 <211> 323  
 <212> DNA  
 <213> Homo sapien

<400> 93  
 tcgatataat agcgaattgg cattaatcat ctgacggcgc agtgtgatgg atcgccgggc 60  
 aggtgtgggc cagcctgta gccccagcta cttgggaagc ttgagacagg agaatcgag 120  
 gaatctagga ggcggagggt gcagtgagcc gagatctcgc cactgcactc cagcctgggc 180  
 gagagagtaa gactctccgt ttctcccaa aaaaaaaaaa aaaaaaaaaa aaactttggg 240  
 gtattattgg tcatgtgttc cctgggtgaa atgggtttcc ggtcaaatcc aaattgataa 300  
 aaataaaaag aaaaagtgac gat 323

<210> 94  
 <211> 625  
 <212> DNA  
 <213> Homo sapien

<400> 94  
 aggaagtcg ggaaaactga tgctatatag ccaatggcta tctgatcagc cgagcggcgc 60  
 aatgtgatgg atgcgtgcgc ggcgaggtag ttctgtggta gtagggcttt gtcacatcat 120  
 gcactaaaaa cagaatgtga ctcaaccttt tctactgctg actgagttgt gatgaggctt 180  
 tttctttcta agaagtgttt aaattaccac atagtccagg aatcacggac agtaacacta 240  
 acactttcat ctgtgtgggc caggagttgg gcatgtagtt taatgacgta taatttttga 300  
 attccaagca tagtttgaaa aaatatgaaa atcttagcac ccagcacatg cctattaatg 360  
 aagaagttct cagcagctgg cagaaatgca tctgtgtaga gagacacagg cggaacaggt 420  
 ggcagggtgg ggcgtcatct ggaggcctgc gtctgggctg agtgaccttc gttcttaggc 480  
 tgctggtgt gggaaacgtg aagatgtgcg catttctccg gcccatgct gggcacttgc 540

52

tgcaggccct tacccttgtc gtttctaaat atcgaacata agaagactgt ccacttctct 600  
 tttaatgtaa ggatgttggg aaacc 625

<210> 95  
 <211> 810  
 <212> DNA  
 <213> Homo sapien

<400> 95  
 aggaagtcg ggaaaactga tgctatatag ccaatggcta tctgatcagc cgagcggcgc 60  
 aatgtgatgg atgcgtgcgc ggcgaggtag tttctgtgga gtagggtctt gtcacatcat 120  
 gcactaaaaa cagaatgtga ctcaaccttt tctactgctg actgagttgt gatgaggctt 180  
 tttcttttcta agaagtgttt aaattaccac atagtccagg aatcacggac agtaacacta 240  
 acactttcat ctgtgtgggc caggagttgg gcatgtagtt taatgacgta taatttttga 300  
 attccaagca tagtttga aaatatgaaa atcttagcac ccagcacatg cctattaatg 360  
 aagaagttct cagcagctgg cagaaatgca tctgtgtaga gagacacagg cggaacagg 420  
 ggcaggggtg ggcgtcatct ggaggcctgc gtctgggctg agtgaccttc gttcttaggc 480  
 tgccctgggtg gggaacgtg aagatgtgag catttctccg gccccatgct gggcacttgc 540  
 tgcaggccct tacccttgtc gtttctaaat atcgaacata agaagactgt ccacttctct 600  
 tttaatgtaa ggatgttggg aaaccaaagc tttatggctt tggaatggaa tttttctcat 660  
 ttcctaaaaa taaatggtag aagtaaagta tgctcatcat gagctgggcc caagcgagt 720  
 tttggtttag ccagaaggta aatgggcaag cagcgtgagc tgacagcttg caaaagagga 780  
 aatgaaaaag gctgtgtac acgttcgcga 810

<210> 96  
 <211> 716  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (590)..(590)  
 <223> a, c, g or t

<400> 96  
 cgggactgat atatataggg gactgggtct tagatgcatg ctcgagcggc gcagtgtgat 60  
 ggatcgagcg gcgcccgggc aggtgtttga gcctaacctg atcaacataa caagacctg 120  
 tctctattaa aattgaaaaa agaaaaagaa taaaagacca atttttttta attataaaa 180  
 ctaattctgc cagctactta tagtcataaa aggtgaatca actaattcaa catgttctct 240  
 ttagtagtca atttttaaaa agcaagtatt aatgggtagt ttaaactctt ctgaatacat 300  
 taccattgta aagaacaatg tttaaaattt acttttcaaa ctaatgcatg cagtttctcc 360

cctttgaaaa acctaacagt attatatgtg gtttagaaca atgtagataa ctttaagcca 420  
 agcaacaaat atttgggcat ttgcatgggc tatgaaataa aatgttgtag taactcttga 480  
 aaaattaaaa aggactgggt ttcttaataa aatataagca tttaatcaaa aaaaaacaaa 540  
 aaaaaacaaa aaacaggcgg gcgggtaact cagtgggcca taggggtggn cccgtgggg 600  
 ggacaatttg gttattcccg gtccacattc accacactac ctcggcacgc gacacaactt 660  
 gaccagcaca gcacaagaga gcaaaacaag caccacagca cacaccagca aaaacg 716

<210> 97  
 <211> 341  
 <212> DNA  
 <213> Homo sapien

<400> 97  
 agcttttttt tttttttttt ttgtgtttt aaatttttaa aaagggttta ttggcagggg 60  
 ggcaggaatt aaacaaaag ggccaaacc catgtgttca tcatcgtgac tcttaagaac 120  
 tctctttttt tctcattttt tcttctctt ctgtggtgca gcaggggcgc aaaaccacgg 180  
 agcaggggcg tggcaaagcc tggggcgagc agacgacggg aacagcccca ccaggcgggt 240  
 accacgggca acgctagggg gacaccatgg gccatcagct ggaccctggg gtggaactcg 300  
 gtaatccggg acacaattcc cacacaacaa cgcgcaagca c 341

<210> 98  
 <211> 903  
 <212> DNA  
 <213> Homo sapien

<400> 98  
 tatcactata tggcaattgt gcctctaata atctcgatgc tggctgcagt gtgattggat 60  
 atgctggcct gccctgggca tgtccccccc cctttttttt tttttttttt tttttttttt 120  
 tttttttttt ttttttttat aaaaaaaaaa aaccggaaa atgggggggg gagggagagt 180  
 gaaaaaaaaa aaaagtgggt gtgaaaagag tgtgtgttcc aaaaaacaag gttgtgttgt 240  
 tatgctcgcc ggagaagaag agagagatgt ttattattgt tgttaggagt ttgtggtggg 300  
 tgtggtagat gagaaccccc actgttgtgt cgtgggttgg catacatatg ttagagagaga 360  
 gctaagaagt atgggtttgt acaaaacaat gatgtttaac cctcctaata ataactaaaa 420  
 acatatatat attatttcca cacacaacaa aaactcgctt tgtccataca acacacacac 480  
 aacaacagaa atcctccacc acaatcagtt atacaaagag tgtgttgtgt atattcatga 540  
 ctgcacacgt cttacaccac acttttcttt tcacaaaaac ttctcccaca tcaaagcact 600  
 ttacttatgt gtgtggcgtg agggctatac atcccttcta ggagaatctc tcgtttaga 660  
 gacaaacgat gtccttctta taccagccc cctcgacagg ccacctgcac gtcttcccaa 720

54

aacacatgac aattatcgtc ccctcctccc acacataaac ctccaagagc attgtcttct 780  
ccccactcct ctgggccac acaatcatac caacacatct aactctctc cccccacaa 840  
ccctcttctc gctccacaac catcatgtcc caaacctcc cccccctt ttccaccact 900  
tcc 903

<210> 99  
<211> 928  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (778)..(778)  
<223> a, c, g or t

<400> 99  
tactatatag gccctgggtc cttagatcat gctcgagcgg cgccagttgt gatggatgcc 60  
gcccgggcag gtaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaagaaaaaa 120  
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 180  
gggggggggt tcgtgccatc ctccccgtc ttctcttct ctattactac ttcccccg 240  
gatcgcgctg cgcgcggggg ggacactcta tattatatag aagagggaga cagacgatac 300  
tctagcaaga gagcagagaa catcgctaag ataggttggc tccccggcga aactattgag 360  
gtcgtcgcca ctatattgga gcccttcgcg tgcgcttg tgcacacaac accacaatga 420  
gtgcagtgt tctattgagt gtggtacca taccgcaggc gcgcataaca cctacctact 480  
gcgccaggcg cgctcactct atgttggtggc agtgcggcg cccgtgtttt ggtatccaac 540  
aggggagggg gggacggcca cactcaatc aacaatcaa tacaccgcac ggcgggctgc 600  
atcttgcgct aacacacatc cttgaggctg ccagcacgac gccgcttct gttccactaa 660  
ctagtgccaa cccgtccgat atatatgaac cgtggcgcg tcgctccgc cactaaagt 720  
gagtgtggtc gatgatcact attataaat acacacacag cgggcgaggg ggggaganga 780  
attgattaaa aaacaccctg ctctgtgtat ttaaccgcgc cgagggtgct agaacaaggg 840  
agggacgaac tatctcatc catccacct gacttgtgga ggaggaggag aacacctctc 900  
cctcttaca taaaaccgcg cgggcggc 928

<210> 100  
<211> 852  
<212> DNA  
<213> Homo sapien

<400> 100  
gccgccccgg caggtacagg acgccccag actgcagccc ttgtcccagg gcaactggtga 60  
gcaacacgca gccatatggc aagtgcctgt gtcctgtcc ttcaggcca tcaattcctg 120

55

```

ggagcttttg ctttatcact ccttcagtct taagtccatc caccagagtc tagaaggcct 180
agactgggcc ccgccatctc gtgcatgaga catgttgact gtgcccggtg ggagatggcc 240
acgctgtgtg tgccaggat atggccctgg agtctgcatt ggcacctgct atagaggcat 300
ttggacggaa tccctcacac catcttctgg tgcctcacgt ttttcccat tactaacaaa 360
atgcataaa cgctgtgaca ttacttaact ctagagttgc cttgcgcagt cgctgtacat 420
tctagagcta ttccaggtag gttgtcaca ttatgtccag agtgaagcat aggtcatata 480
agcctaaggt tccatcctgg gggattccag ctagggcgtc ctgaggagaa ttgcgagatc 540
acacatcaca ctctgtggga tctcagggat agcgatgtcc cgttcccat gccccagct 600
aggcttcaca ggaaccacag ttgcgcagt cctgcaagct ttaagtgaca gtcggtgtcc 660
tgaaaagccc cagcaagttg cccaggtac ctgggaagac cacgggatct cttttactac 720
ccacgatgac tccggggttt ctgggcaagg ggccaggagg cacatggatc cctctgcagc 780
acatccgcc gttcaagttc gtccaacaat gcaggccttt ttgtaaacac aaatgggcc 840
ggcacgccgg aa 852

```

<210> 101  
<211> 254  
<212> DNA  
<213> Homo sapien

```

<400> 101
gatgaataaa ctacattggc aatggcctct atcatctcga cggcgccagt tgatggattt 60
tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 120
tttttttttt tttttttttt ttttgggggg ggggacaggg gagcaggggg ggcgcgcggg 180
gggagaatgt gttctcccc cccaccccc ccaaaaaaaaa aaaaaaaga attcgataaa 240
taaaaaaaaa aagt 254

```

<210> 102  
<211> 447  
<212> DNA  
<213> Homo sapien

```

<400> 102
tcgcggccga ggtgggaggt ctaggctgca gtgagccggg acgatgccac tgcactccag 60
cctgggcaac agagtgagac cctgtccag cactctggga ggcagaggag cccagttgga 120
gatcagcctg ggtaatatag tgaaacttga tctctacaaa aaaaagaaga aaaaaaaaaa 180
ccgcgtgtgg tggcgcgcac ctgtagtccc agctactggg aagctgaggt gggaggatca 240
cttaagccca ggaggcagag gtcacaatga gccgaaattg tgccaactgg actccagcct 300
ggggcaacag aggaaggaac tcttcaccag gaaaaaaaa aaaaacaaaa aaaaaaaaaa 360

```

56

aggcgggggg ggaaacacag gggcccaaac gcgggggaccc ggggggggaa atgggggaac 420  
 ccgggaccac aaattcccaa aacaaag 447

<210> 103  
 <211> 697  
 <212> DNA  
 <213> Homo sapien

<400> 103  
 gcgtggtcgc ggccgaggtc tccctttttt tttttttttt tttttttttt tcatttttta 60  
 aaaaaagtaa ctggttttta taattatggg aaggtggggc cggattaagg gggtttagtt 120  
 gttgcctcag ggaattgggt gtggacgtgt gaaaattaat taaaaaaaag gctgtgaaag 180  
 aaaaggggtg tggttttgaa ggccaggcca aagggttttc ttctaggctc cgtttcgtgg 240  
 aaaggaacag cctatttaga aaggattatt ggacaacgcc acattactat aggccccac 300  
 aatctcacat atttaaaaaa tttccgtaga aacaacttat agctctgaat ctactaccg 360  
 tgggtgggtg tctccacgtt tctcttctaa atacagtgcc ggactcagag gaaccccccg 420  
 aggggtctcc ttgctggtt tcttttggtg taaaaggaca ggctatagtc ttcgtgtata 480  
 ttctcacata aagcctgtgg gggatacatc cagagggtca caaataaggt ggtatacacg 540  
 ccgggtggct aaacaagtgg gtcactcgc gccctcaca atattcacca ccacaacaat 600  
 accccacgca cacaacaccc atcaaaaacc acaggggggc aggaaaagac gcccaaccaca 660  
 gacgaaaaca aaaagagcag ggaaaaaaa caaaact 697

<210> 104  
 <211> 807  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (380)..(380)  
 <223> a, c, g or t

<220>  
 <221> misc\_feature  
 <222> (404)..(404)  
 <223> a, c, g or t

<220>  
 <221> misc\_feature  
 <222> (618)..(618)  
 <223> a, c, g or t

<400> 104  
 gcggcgccc gggcaggtac cacactaagt ctctgggcc ttgtgacttc ctgtgaggat 60  
 gtgtggtgag ggccaaagtg ctatggtttc ctgctccag tgatagatgg agataaagtg 120



57

```

cttctcatgg ccccgctccaa tgccctgggtg aaggactgtg gcactccaaa gcgtgagcca 180
gaggggtaat ctgcctgatg tctcgtecca ttcaatctcc tgctggaccg ttggggaggca 240
ttctagagct ctatgctgtg gcacgtggac atccctcatg agcaagactc ctcgtagacc 300
ataagtacg attgtagcat tccctgataa gcgcgtctat gcattgactc caattctatc 360
tccattttcta gaggtagctn tgtgtggcac accattttctg tccncatttc agctgttcag 420
ctacatctta gctcgagttc tatctaaacg ctcgcttttg cctttgggtg gactcgatat 480
agtttgggtt tattgggcgt tgtgcaaact cactatgctg cagcttgata tctttaccag 540
ttggcgcaag aaacgaacac cttggcagga ctttcttttt cccatttcat tcatgacttg 600
tgccaattg tggcccanca agggctctat gcattctaaa ccattccttg aaggcctttc 660
cttccaagtg gagcttcccg ttgtggaagg ccacattgtc gtgggggcac ccttgggttg 720
cctgtgtggg cccacgcttg gcttctttgt tgccttgaa cgtgtgcctt cccggtcctt 780
cggggaggaa tttctttggt cccttgg 807

```

```

<210> 105
<211> 975
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (548)..(548)
<223> a, c, g or t

```

```

<220>
<221> misc_feature
<222> (572)..(572)
<223> a, c, g or t

```

```

<220>
<221> misc_feature
<222> (786)..(786)
<223> a, c, g or t

```

```

<400> 105
cagggagatg tccctggggc agacactaag gcagggtgtg aagacaagct gcttgtcaag 60
aagcatttcc cggcaagaga ggggcaagtc tggggctcca actgggtaca gcctgggtgc 120
agttataagc ccccttggct tacttggtag aagatggcta cttggatgta cctcacttaa 180
agatgttttg taccacacta ggtctctggg cccttgtgct tctgtgggt ggggtgaggg 240
ccaaagtgct atggtttctt gcctccagtg atagatggag ataaagtgct tctcatggcc 300
ccgtccaatg cctgggtgaa ggactgtggc actccaaagc gtgagccaga ggggtaatct 360
gcctgatgtc tcgtccatt caatctctg ctggaccgtt gggaggcatt ctgagctct 420

```

```

atgctgtggc acgtggacat ccctcatgag caagactcct cgtagaccat aagtgacgat   480
tgtagcattc cttgataagc gcgtctatgc attgactcca attctatctc cttttctaga   540
gttgcgnttg tgtggcacac cttttctgtc cncatttcag ctgttcagct acatcttagc   600
tcgagttcta tctaaacgct cgcttttgcc tttgggtgga ctcgatatag tttgggttta   660
ttgggcgttg tgcaaaactca ctatgctgca gcttgatata tttaccagtt ggcgcaagaa   720
acgaacacct tggcaggact ttctttttcc catttcattc atgacttggt gccaatgtg   780
gccancaaag ggctctatgc attctaaacc attccttgaa ggcctttcct tccaagtgga   840
gcttcccggt gtggaaggcc acattgtcgt gggggcacc ttgggttgcc tgtgtgggcc   900
ccacgttggc ttctttgttg ccttgaaccg tgtgccttcc cggtccttcg gggaggaatt   960
tctttggtcc cttgg                                     975

```

```

<210> 106
<211> 735
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (627)..(627)
<223> a, c, g or t

```

```

<400> 106
gcgccgccc gggcagggtgc tttttttttt tttttttttt ttttttgggg gggtaaacttt   60
tttataaacc ccccagcatc cttacacaaa aacctaccaa tgtgggaacc ctttcaccaa   120
atctccgtga ggaatgtgtg ctcatatata taaaaatgtg tttaaaagggt attgtgtaac   180
cattttattct tctccatata tgtgtatgtg cgcaacaatg tgcacaaaac gccatagtgt   240
gtgctccact cgtgttataa gttctaacag caccgccact ataagacagg gagaaatact   300
tctctctcca caaagggttt cacttttcca caaaatataa ggtgtgacag ggcgcgccac   360
agtgtgtgtg tgcggtgtct tttgtgagag aggtcgtgcg caccagtgtg tgtggagaaa   420
gagactctcc acagactata aaacatgtag acaccactct ctgtgtgtac cccacactc   480
tctctctcag agagaacctt ctctttctca caaagegtct gtgagcggcg cgtccccaca   540
cacaagaga gagagagcag agaagacgct ctatttattt ctctgagcca acacacggcg   600
tgcggagatt tgtgcgtctc ctctgngct ctctcgaggg ggctcctctg tgtggactct   660
ctgagcttat aaaatgttgt gcgtcccacc atctcggttt tctctctca tttgaggaaa   720
gagcttgggg gggaa                                     735

```

```

<210> 107
<211> 751

```

59

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 107

```

gcgtggtcgc ggccgaggat acccgtgccc agtgaggacg ccgagctcca gccccgagcc      60
ctggacatct actcgtgcca gtggatgatg ccttcccacg agcaaggagc tgatcgaagg      120
tcgctgtaaa ggaatgtctt gaagaaaggc tcaagagtaa acgtgattcc tccattctat      180
gaggaaatgaa gtatggtcca agatcccat ggtgatgact gccgtgttgc agcagttgtg      240
tccgatgctg tagtgaaaag gggtcggagg atcgggtaag gctgtgtgac tgtctcctcg      300
agtgagcctc catgctaatt cccttccctc gcttgaaata gtgcttgta gtggaagggtg      360
gtgctgggtc gaatatctcg ctacatact gtcgaccac catcctcgtc ttacggttgc      420
ccacaatgaa ggtaccaaca atcttttcac ttcacacatg agaagttatg gcattaagca      480
aacaagatca aagtgtttgt attttccgtc tgaacgggga gaacggggcg tccgttttgt      540
cccctgggcy tggtttcccc agaacacata aacacagaaa accaacaatt taggaattgg      600
tcccaaaaca acaaacaaga gcaaacagag aagagaaaac aaaagaggcg cgggcgggta      660
acaccccgty ggcccaacga ggggtgtccc gcgggggtgg aacaggtggc tcccgcgccc      720
acaattcccc accaacacgy ggccacaacg g

```

&lt;210&gt; 108

&lt;211&gt; 640

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 108

```

cgccagttat gatggccgcy cgggcaggty gggcaggtaa aaaaaaaaaa aaaaaaaaaa      60
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa      120
aaaaaaaaaa aaaaaaaaaa gggggggggg gttttttttt ctcccccgcy tgagtggcgc      180
ccccctctt ctttttctt ctctggttg tttgttctc ttttattat tatgataata      240
ttatgtctta ttaatcataa tattatgtgt tgggggggtg ctcttcgcy tgattatcta      300
tcaatatctg tttgtgtggt acagatttct agccgcggtg tgtctccctg cgcgcgtgat      360
aaaacaacag ccctctctct cctctccgty tcttctctt cttatttgty ctaatccagc      420
aaacgaagag aaagatgcaa cacacttgt tggctcagty tcctgactcy aaccatcgca      480
cccagcgaay caaaaacaga agaacagaga cggtcgggcy gggacagtaa tgctagtggg      540
caacaatgta ccccccgcy ggtgagacaa gaaactatcy tcttctacgy ccgcatgaac      600
ttctaccaca actaaacaaa tgacgcaaca aaaaaaggcy

```

&lt;210&gt; 109

&lt;211&gt; 533

&lt;212&gt; DNA

60

&lt;213&gt; Homo sapien

&lt;400&gt; 109

gagcggcgcc gtgtgatgga tggtagacagaa ggaaaacaac tttttatgta tacttctaaa 60

aggggaaaaa aaaaaaaaaa gagaaaccct ttgatttcca cgttgcccat tcgtcaagac 120

atttccactt cacagatttt gaggtttctg atttccaggt tctgagtttt cccaattggt 180

taattgttaa ccagaacttg gcacacacac atttaagaat gaattgttaa tttattttatt 240

tcctctttgc tggtcattac cgtcgctttc tattttcttc ttttcttttg tgttgaattt 300

tattttataa gaacaaaaaa cttttttgct aacgacttat ttgacagttt taaaaattca 360

attaaccccc gtttttttca ggaaacaaaa aaagaaaaaa aaaaaaaaaa aaaaaaaaaa 420

aaccctgtgg tatatatctg tggccaaata gccttttctc cgtgggtgtg ttaaattggt 480

taactccgca catcaaaatt cccacaaaac tatatgtgac acacaaaggg agt 533

&lt;210&gt; 110

&lt;211&gt; 262

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 110

tgtaacaat aaggcacgcy ttttgctttg gtcgcttatt atccactac gagactacta 60

cagagccaag tacctgagcc actgcgcgca ggggactcgg gaatgtctcc atggtcacaac 120

gaacgcagta ttgccaata tctcatggac aaagtgacaa cagcactaca agcaacaat 180

cacataagcc catacatcga tcaacaaaga tactacaact acgccagcgt agggatacaa 240

cccagactga ctcacatcac aa 262

&lt;210&gt; 111

&lt;211&gt; 1494

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 111

tgacaggtac aggatatagc ctggcacttt cctgtagtct acacacaatg cccaactgcc 60

tgaccttagt ggtagtgtct agagtgatct cctgtccatc agcacaggac agtcagaatc 120

tcatectttc atgcggccaa catccccaga ccctttatgt tgacgccagg acctcatctc 180

acctctccat cctcacctta caccgcccct gcctgaccag acaaccaccg gagcaccagt 240

ttggattcta ccgaaaccac ctactcgtca cttctgttac ccaccactat cttgactgac 300

tgacacaccc eggcatcac ctacttatac ttatactatt atgactatga atactcgttc 360

ttaccttacc acctttggat cactacactc ttactcctcc ccacaactgt ggtgtgacac 420

actgacactg gtacgccacg gttcgtccct cggtcacaac acacgcaccg accctaccgc 480

ttatecttca ccctactgcc cttacctcgc cgaacacttc acacttctgc acaaactatc 540

61

```

ctcgatgacc cctggacgcc tggacatggc gatgccctac gttctcgcac cacaccttgc 600
aacaccgact cccccctcac tcacaccact acgaaacaac accacccctt cgcaccacca 660
caccataact taccttataa ccgcccccta ccacagaacc ctactaactt cccaacaca 720
cccctacggc gatgaccacc ttacctata cctaacctta acaacccctc tgaacctcg 780
acccacacac cgttacccat taccaccatt aaaccctc cggattacaa cccaacacac 840
atccgacggc actacgccct tcaggaacac ccacctaaa ctacaccac tctattatac 900
aaccaacac cactactact atgcacacca caaccaacc caaacatcca ctaccacat 960
aaaacactca gcaggacaac actctgagca acaacagtga ctggacacga cccgcagaac 1020
acacacacac ccacacgctg aggagagaaa caacaaaaca cgccaccccc ctgtcaagcc 1080
accacgaaac caccggaacg atggccgcac accaacacc gacagaagcg agtcataaac 1140
ccctaatecc gctcccaaca cccacccgaa ctaccaaccg acctgcgcag aacgctcaac 1200
ggcaagtaac atcacagagc tgactgctcg ttctccctg atgcggtgac gatcgagccg 1260
tagcctacgc gtctccagt cgcgcacgag gggcgacggg ctcggtcgcg gcagtcgtgg 1320
caatgaatcg gccagacgag ctcgagccg cgcgacggac cagggacggg gtgagcgtgt 1380
cggcacgcag ctgtcgacat catcatcac tctcttctt ccgcgttcg tggcgggcgg 1440
gaggaccgcg ctgcagactg gtactctgag ccaggctagc cgacctcacc ccgg 1494

```

```

<210> 112
<211> 811
<212> DNA
<213> Homo sapien

```

```

<400> 112
aggagtggaa tcatattggg cgacctgggc ttatagatgc atgctcgagc ggcgagtggt 60
gatggatcgg ccgcggggca ggtcctcccc tttttttttt tttttttttt tttttttttt 120
ttaagagggc caaatcttga tccctttttt gtaaaaaaaa tttttttttt tttttttttt 180
ttttttttgt ggaaaccccc tttagaaacc agtgctgcgg ccctcccagt cagcatgtgt 240
ctgttgcgc gccactcttg tgtatacaa agggatgggt cccagcagg gtggaagagg 300
gagtggccac cagtgccgg acgaggggtga caccacgcg gcgttacaca ttctttggaa 360
acaccacgc gtgggtctcc cgggtatat aaaactcctc ccccccccta tagagtgtgg 420
cgacatctgc gatatctccc cgcgcggggg cgggtgctgt cccaccagt tgggtctcct 480
cgagggcccc cacaggacct cctcaggtgt gcgtcctccc ctltattaga gggtagggca 540
caacaccac cccccctcg agtcgtgcgc ggggacaacc ctctgtagcg gaccacgaa 600
ccaccagaaa agtcctatct ctacgcgcg cgcgaggaac cctccgcgag ggccgcggac 660
aactgcaagg gatatttccg cgcgccaca caccgtgggg gggcaccaac cgcggggccc 720

```

62

aaacagcgat gttaccgcgg ggtggcgaaa attgtgtttt ccccgccctc aaaatctccc 780

ccaccacaaa ctaccacca ccccaaccag g 811

<210> 113

<211> 1506

<212> DNA

<213> Homo sapien

<400> 113

tggctctgctg gcctgaggte ccccccttt tttttttttt tttttttttt tttttttttt 60

tttttttttt gaggggtggc cggggggggc aagagagagt gtgtgtgcct atatactagg 120

tgtgggtggga gagagtgtg gagagtggg gtgtataaaa atgtgtttat tttgtggtgt 180

gtgtgtgtgc tcactaatag agaggtggag gtggtgtgag aatataaacc aactggaag 240

tgtgtgaatg aatataaaca gcctatatat tctcgccgcg aacagcgcgg tgtgtgtata 300

tatgagagaa gtggtgttag agagagtggg gtgtggcggg tgtgggtgca cactgctgcg 360

ctgcggcggg ggtgttctct ctctctcacg agctgtgtga tgatgaacac acaagagta 420

ggtattatat attctctcct aacgcgcct ctctctcgc gcgcgcgcgc aaaaacagag 480

gtgggacaat agagagtgtg tgctatagcg cgcgtgcaaa cacacaaaat atatacagag 540

agatgtgtgt acaaccatat gacacaaaca cacagatgaa caacaacat atttttgcaa 600

acaaaaaaca gctgtgtaat ataagagtgt gtgtgtgtgt gttccctgc gagagtattt 660

acatatatat ctctcccacg cgcgaggggac aacacacatc ttttaccata gagagatgag 720

tgccccccca gggttatata acacacacaa acgcgtgctc tccgcggagg gagacaaac 780

aacatatcta ctgtgtggag agaaaaaaat ataacttctc tacacctttt tgagcagaaa 840

cacctgtgtg cgggctatac acatcacgac gggggggcgac aaaaaaatg gtgtacacc 900

ccctggggtg tgcgaaaaa acatgctgtg ctacacacacc gccgcggtct ccaaaaaaat 960

tctccccaca acaccaacac ctccagatc aaagaccacc acacaacaat gagtgcgata 1020

ctcacagcac ttcacgtaca tctcagctg acgccattca tccaccaaat caatactgcc 1080

tcgaacttat actctacat tctcttagc acctcactgc caegaacacc actctccctg 1140

aacacagaca ttcagtcac acctatcaca aaccaaata catcccacc gctcaccatc 1200

tccactactc tacataaaca caaacctcac tccccacaa ccaccacaca cacctactac 1260

atccaaccac acaacactcc cagcacctc aacttcacca ctctctcact acaaaccttc 1320

tcacacatca cgccacacat ataccacacc tctcactcaa ccaaccacaa aaacaaacaa 1380

actacaccac actccacat ccccaaccaa actcccacaa ccaacaaaa tcacaacaca 1440

cacccactc acaccaacac acacaccacc acaccccccc ctttaccacaa tacactctaa 1500

aaacac 1506

63

<210> 114  
 <211> 779  
 <212> DNA  
 <213> Homo sapien

<400> 114  
 aaaaaacaaa aaaaaacaaa aagaaagagg aatgaataat cactataggg gcctcggtgt 60  
 atctagatgc atgctcgagc ggcgcattgt gatggatcgt ggtcgcggcg aggtgcttat 120  
 tttttttttt ttttttttgg tccatgttta aaaaaagtgg aactatgggc ttaattatca 180  
 atggggccagg gggggcctga ataagggggt tagtcgtgct caaggggatg ggtgtgggcg 240  
 ctggtggaag atagatcgac aaaaatgtgc ttgaaatgag aaatgggtgt gttggtgtta 300  
 agaaggtgcc atgtgcccaa tgggtgctcc tcattgtgctc tgcattctctg ggagaatgag 360  
 cgacacgcct ttgagagaaa gagatgtcat tggcaacgcc atggtatcag gcgcccacca 420  
 aatcaatata ttacaacaaa tatctctgga aaacattctca cgtctggacc atccactggt 480  
 cgggtgtgtc catgttcttc ccatcaatgc gcggtcagtg gaccaccaag gagtccttct 540  
 gggctccttg gtaagaagcg cagctaagtc ctgtgttata ccatagaatg tctgggctgt 600  
 aaatctatgg gcacattaac gctggtatcc ctggtgtgga gacaattggt cacatcgcg 660  
 tcccaacata ttcccaaac aaaactatac agagaaccaa gagacaaaaa taattggaaa 720  
 gggcacacaa gacaacaacg gaacccaaaa aaaagcaaga aaaaacaaca gggacaaca 779

<210> 115  
 <211> 195  
 <212> DNA  
 <213> Homo sapien

<400> 115  
 tgctctgtgt ctgttctgtg ctgctgtgct gatgctgtgt atcatgctcc actcaaatgt 60  
 gctgtgtcaa tactgtgtct atccacatga catcatgggt gattaactgc atgtgaaatg 120  
 aacattgttg agcaaaatgt gccatgcaaa atgtgccagt gaacctgtaa aaatgtgcct 180  
 gctgtttgct tggct 195

<210> 116  
 <211> 62  
 <212> PRT  
 <213> Homo sapien

<400> 116  
 Met Pro Ser Gln Asn Ala Val Phe Ser Gln Glu Gly Asn Met Glu Glu  
 1 5 10 15  
 Glu Glu Met Asn Asp Gly Ser Gln Met Val Arg Ser Gln Glu Ser Leu  
 20 25 30

64

Thr Phe Gln Asp Arg Gly Arg Gly Leu His Gln Arg Gly Val Gly Pro  
 35 40 45

Ala Val Pro Ala Arg Ala Ala Asp Pro Ser Tyr Cys Arg Pro  
 50 55 60

<210> 117  
 <211> 414  
 <212> PRT  
 <213> Homo sapien

<400> 117

Gln Glu Ser Leu Thr Phe Gln Asp Val Ala Val Asp Phe Thr Arg Glu  
 1 5 10 15

Glu Trp Asp Gln Leu Tyr Pro Ala Gln Lys Asn Leu Tyr Arg Asp Val  
 20 25 30

Met Leu Glu Asn Tyr Arg Asn Leu Val Ala Leu Gly Tyr Gln Leu Cys  
 35 40 45

Lys Pro Glu Val Ile Ala Gln Leu Glu Leu Glu Glu Glu Trp Val Ile  
 50 55 60

Glu Arg Asp Ser Leu Leu Asp Thr His Pro Asp Gly Glu Asn Arg Pro  
 65 70 75 80

Glu Ile Lys Lys Ser Thr Thr Ser Gln Asn Ile Ser Asp Glu Asn Gln  
 85 90 95

Thr His Glu Met Ile Met Glu Arg Leu Ala Gly Asp Ser Phe Trp Tyr  
 100 105 110

Ser Ile Leu Gly Gly Leu Trp Asp Phe Asp Tyr His Pro Glu Phe Asn  
 115 120 125

Gln Glu Asn His Lys Arg Tyr Leu Gly Gln Val Thr Leu Thr His Lys  
 130 135 140

Lys Ile Thr Gln Glu Arg Ser Leu Glu Cys Asn Lys Phe Ala Glu Asn  
 145 150 155 160

Cys Asn Leu Asn Ser Asn Leu Met Gln Gln Arg Ile Pro Ser Ile Lys  
 165 170 175

Ile Pro Leu Asn Ser Asp Thr Gln Gly Asn Ser Ile Lys His Asn Ser  
 180 185 190



65

Asp Leu Ile Tyr Tyr Gln Gly Asn Tyr Val Arg Glu Thr Pro Tyr Glu  
 195 200 205

Tyr Ser Glu Cys Gly Lys Ile Phe Asn Gln His Ile Leu Leu Thr Asp  
 210 215 220

His Ile His Thr Ala Glu Lys Pro Ser Glu Cys Gly Lys Ala Phe Ser  
 225 230 235 240

His Thr Ser Ser Leu Ser Gln Pro Gln Met Leu Leu Thr Gly Glu Lys  
 245 250 255

Pro Tyr Lys Cys Asp Glu Cys Gly Lys Arg Phe Ser Gln Arg Ile His  
 260 265 270

Leu Ile Gln His Gln Arg Ile His Thr Gly Glu Lys Pro Phe Ile Cys  
 275 280 285

Asn Gly Cys Gly Lys Ala Phe Arg Gln His Ser Ser Phe Thr Gln His  
 290 295 300

Leu Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Asn Gln Cys Gly  
 305 310 315 320

Lys Ala Phe Ser Arg Ile Thr Ser Leu Thr Glu His His Arg Leu His  
 325 330 335

Thr Gly Glu Lys Pro Tyr Glu Cys Gly Phe Cys Gly Lys Ala Phe Ser  
 340 345 350

Gln Arg Thr His Leu Asn Gln His Glu Arg Thr His Thr Gly Glu Lys  
 355 360 365

Pro Tyr Lys Cys Asn Glu Cys Gly Lys Ala Phe Ser Gln Ser Ala His  
 370 375 380

Leu Asn Gln His Arg Lys Ile His Thr Arg Glu Lys Leu Cys Glu Tyr  
 385 390 395 400

Lys Cys Glu Gln Thr Val Arg His Ser Pro Ser Phe Ser Ser  
 405 410

<210> 118  
 <211> 160  
 <212> PRT  
 <213> Homo sapien  
 <400> 118

66

Met Gln Leu Val Leu Leu Val Pro Val Cys Pro Thr Ile Gly Val Phe  
 1 5 10 15

Phe Arg Arg Leu Gly Pro His Phe Asp Val Gly Arg Phe Leu Cys Leu  
 20 25 30

Trp Gln Phe Val Val Pro Gln Ser Leu Pro Cys Arg Trp Arg Gly Ala  
 35 40 45

Arg Gly Phe Arg Thr Leu Gly Val Leu Phe Leu Val Val Pro His His  
 50 55 60

Gly Ala Ser Ser Gly Cys Arg Leu Arg Arg Cys Arg Phe Phe Cys Ser  
 65 70 75 80

Cys Gly Ser Ala Ser Val Asp Leu Phe Ala Leu Gly Trp Ile Cys Leu  
 85 90 95

Ser Leu Arg Arg Pro Ser Val Arg Cys Arg Trp Ile Pro Leu Val Thr  
 100 105 110

Ala Arg Val Ala Cys Ala Ala Cys His Ala Gly Thr Pro Pro Leu Cys  
 115 120 125

Ala Phe Leu Gly Arg Cys Ser Ile Thr Ala Cys Cys Thr Ser Phe Cys  
 130 135 140

Phe Ser Leu Phe Thr Ala Phe Val Cys Pro Val Ala Cys Met His Arg  
 145 150 155 160

<210> 119

<211> 121

<212> PRT

<213> Homo sapien

<400> 119

Met Arg Glu Lys His Asn Arg Arg Arg Gln Gln Pro Asp Glu Asp Thr  
 1 5 10 15

Gln Arg Glu Ser Lys Lys Pro Gln Gln Ser Ser Thr Lys Thr Thr Gln  
 20 25 30

Thr His Lys Val Ile Pro Tyr His His Asp His Ser Pro Thr Thr Gln  
 35 40 45

His Arg Lys Asp Lys Asn Val Lys Ala Arg Asp Gln Pro His Pro Asn  
 50 55 60

67

Ile Ala Glu Asn Asp Glu Thr Pro Gln Lys Val Asn Asn Met Met Lys  
65 70 75 80

Asp Lys His Asn Lys Ala Lys Pro Asn Thr Lys Gln Ala Lys Lys Gly  
85 90 95

Lys Lys Asn Arg His Asp Ser Asp Ser Arg Ser Thr Lys Arg Ile Arg  
100 105 110

Arg Lys Gln Ile Lys Thr Thr Asp Arg  
115 120

<210> 120  
<211> 15  
<212> PRT  
<213> Homo sapien

<400> 120

Met Trp Ala Thr Val Val Leu Leu Arg Gln Lys Lys Lys Arg Thr  
1 5 10 15

<210> 121  
<211> 97  
<212> PRT  
<213> Homo sapien

<400> 121

Met Lys Lys Glu Ile Phe Pro Leu Phe Ser Asn Arg Pro Ser Ser Pro  
1 5 10 15

Thr His Glu Ser Tyr Pro His Leu Leu Leu Leu Pro Val Arg Lys Tyr  
20 25 30

Gly Ser Cys His Thr His Pro Asp Ala Ser Val Leu Pro Pro His Cys  
35 40 45

Leu Ser Asn Val Ser Leu Ser Leu Gln Cys Phe Asp Arg Lys Gly Gln  
50 55 60

Arg Thr Leu Gly Ser Gly Thr Arg Val Phe Thr Leu Gln Ala Leu Met  
65 70 75 80

Glu Phe Glu Gln Asn Pro Ala Ser Phe Ile Thr Val Arg Ser Gly Trp  
85 90 95

His

68

<210> 122  
 <211> 19  
 <212> PRT  
 <213> Homo sapien

<400> 122

Met Glu Thr His Leu Glu Ala Phe Pro Trp Gln Ser Val Thr Arg Ile  
 1 5 10 15

Pro Asn Leu

<210> 123  
 <211> 59  
 <212> PRT  
 <213> Homo sapien

<400> 123

Met Ser Val Thr Phe Thr Cys Gly His Leu Tyr Lys Gln Cys Ser Phe  
 1 5 10 15

Asn Ser Asn Gly Ala Leu Thr Tyr Gly Gly Gly Lys Lys Thr Thr Arg  
 20 25 30

Ser Asn Trp Ser Cys Gly Asn Asn Asn Ser Pro Leu Leu Leu Asn His  
 35 40 45

Pro Tyr Ala Ala Gly His Val Leu Arg Ala Pro  
 50 55

<210> 124  
 <211> 41  
 <212> PRT  
 <213> Homo sapien

<400> 124

Met Ala Ala Ala Met Ser Pro Ile Pro Leu Ala Phe Ser Asp Leu Ala  
 1 5 10 15

Thr Ser Ser Ser Arg Gly Arg Val Ser Tyr His Pro Ala Leu His Leu  
 20 25 30

Gly Ser Pro Cys Asp Tyr Phe Asp Gln  
 35 40

<210> 125  
 <211> 84  
 <212> PRT  
 <213> Homo sapien

<400> 125

69

Met Gly Gln Arg Leu Leu Val Leu Phe Arg Cys Pro Gly Ala Arg Thr  
 1 5 10 15

Val Cys Thr Ser Ser Thr Glu Ser Gln Phe Gln Pro Asp Leu Leu Lys  
 20 25 30

Cys Val Thr Lys Gly Val Ala Glu Phe Glu His Ile Ala Tyr Leu Lys  
 35 40 45

Leu Gln Ile Ala Thr Met Trp Val Ser Lys Leu Asp Tyr Phe Cys Leu  
 50 55 60

Tyr Gly Thr Ala Leu Thr His Ser Pro Ser Trp Ser Ser Gln Leu Gly  
 65 70 75 80

His Ser Cys Leu

<210> 126  
 <211> 28  
 <212> PRT  
 <213> Homo sapien

<400> 126

Met Leu Phe Phe Lys Lys Leu Thr Leu Phe Asn Asn Tyr Asn Asp Thr  
 1 5 10 15

Glu Arg Cys Pro Ser His Thr Glu Ser Ser Arg Phe  
 20 25

<210> 127  
 <211> 23  
 <212> PRT  
 <213> Homo sapien

<400> 127

Met Trp Gly Tyr Leu Pro Ala Leu His Gln Phe Ser His His Asn Leu  
 1 5 10 15

Ser Pro Gly Asn Lys Gln Arg  
 20

<210> 128  
 <211> 38  
 <212> PRT  
 <213> Homo sapien

<400> 128

Met Gln Ile Met Ile Leu Val Thr Ile Leu Leu Thr Leu Lys Thr Glu

70

1 5 10 15

Leu Ser Asp Thr Pro Phe Arg His Gln Thr Gly Tyr Glu Val Ala His  
20 25 30

Thr Trp Asn Arg Pro Lys  
35

<210> 129  
<211> 55  
<212> PRT  
<213> Homo sapien

<400> 129

Met Ser Gln Gly Gly Tyr Cys Pro Ser Cys Phe Gln Ser Leu Ser Lys  
1 5 10 15

Arg Leu Gly Ala Arg Lys Arg Val Phe Val Leu Leu Asn Val Ser Asn  
20 25 30

Glu Cys Thr Val Glu Ala His Gly Glu Ser Leu Arg Trp Arg Glu Lys  
35 40 45

Ser Gln Lys Gly Arg Leu Leu  
50 55

<210> 130  
<211> 171  
<212> PRT  
<213> Homo sapien

<400> 130

Met Ala Lys Phe Val Ile Arg Pro Ala Thr Ala Ala Asp Cys Ser Asp  
1 5 10 15

Ile Leu Arg Leu Ile Lys Glu Leu Ala Lys Tyr Glu Tyr Met Glu Glu  
20 25 30

Gln Val Ile Leu Thr Glu Lys Asp Leu Leu Glu Asp Gly Phe Gly Glu  
35 40 45

His Pro Phe Tyr His Cys Leu Val Ala Glu Val Pro Lys Glu His Trp  
50 55 60

Thr Pro Glu Gly His Ser Ile Val Gly Phe Ala Met Tyr Tyr Phe Thr  
65 70 75 80

Tyr Asp Pro Trp Ile Gly Lys Leu Leu Tyr Leu Glu Asp Phe Phe Val  
85 90 95

71

Met Ser Asp Tyr Arg Gly Phe Gly Ile Gly Ser Glu Ile Leu Lys Asn  
                   100                  105                  110

Leu Ser Gln Val Ala Met Arg Cys Arg Cys Ser Ser Met His Phe Leu  
                   115                  120                  125

Val Ala Glu Trp Asn Glu Pro Ser Ile Asn Phe Tyr Lys Arg Arg Gly  
                   130                  135                  140

Ala Ser Asp Leu Ser Ser Glu Glu Gly Trp Arg Leu Phe Lys Ile Asp  
                   145                  150                  155                  160

Lys Glu Tyr Leu Leu Lys Met Ala Thr Glu Glu  
                   165                  170

<210> 131  
 <211> 15  
 <212> PRT  
 <213> Homo sapien

<400> 131

Met Leu Ser Arg Ser Val Ala Arg Leu Glu Cys Ser Gly Thr Ile  
                   1                  5                  10                  15

<210> 132  
 <211> 51  
 <212> PRT  
 <213> Homo sapien

<400> 132

Met Leu Phe Leu Gln Met Pro Cys Leu Phe Arg Val Cys Ser Gln Met  
                   1                  5                  10                  15

Leu Pro Glu Gly Glu Thr Phe Phe Leu Cys Gln Ser Arg Phe Leu Gln  
                   20                  25                  30

Ser Ser Ile Thr Pro Gln Lys Val Arg Ser Lys Arg Arg Leu Thr Phe  
                   35                  40                  45

Ser Asp Lys  
                   50

<210> 133  
 <211> 60  
 <212> PRT  
 <213> Homo sapien

<400> 133

72

Met Cys Val Cys Pro Val Pro Val Tyr Gln Leu Thr Asn Trp Glu Thr  
 1 5 10 15

Pro Arg Pro Trp Asp Pro Arg Thr Ser Asn Ser Val Ser Gly Met Phe  
 20 25 30

Leu Arg Trp Ala Arg Gly Ser Pro Arg Val Phe Phe Phe Phe Phe Phe  
 35 40 45

Phe Leu Leu Glu Ala Ile His Lys Lys Leu Phe Ser  
 50 55 60

&lt;210&gt; 134

&lt;211&gt; 32

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 134

Met Phe Pro Gly Asp Phe Ser Ala Phe Lys Leu Leu Glu Thr Ala Glu  
 1 5 10 15

Ile Phe Val Lys Ser Lys Leu Phe Trp Lys Asn Glu Leu Ala Cys Ser  
 20 25 30

&lt;210&gt; 135

&lt;211&gt; 136

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 135

Met Phe Pro Arg Ile Leu Phe Ser Tyr Tyr Pro Ala Leu Tyr Phe Phe  
 1 5 10 15

Val Asn Thr Pro Pro Thr Arg Ile Phe Phe Thr Ser Asp Asn Arg Gly  
 20 25 30

Gly Pro Leu Gln Ile Leu Phe Thr Lys Trp Gly Thr Asn Gly Glu Asn  
 35 40 45

Lys His Arg Trp Val Trp Val Glu Leu Asn Arg Ser Thr Thr Ser Gly  
 50 55 60

Gly Leu Ser Ser Glu Lys Arg His Thr Thr Ser Gly Glu Gly Ala Ser  
 65 70 75 80

Pro Pro His Pro Glu Asn Ser Pro Arg Ala Phe Arg Pro Arg Arg His  
 85 90 95

Leu Val Val Ala Leu Arg Arg Ala Pro Pro Pro Phe Phe Phe Phe Phe



73

100

105

110

Phe Phe Phe Phe Val Phe Phe Phe Phe Phe Phe Phe Phe Phe Leu Ile  
 115 120 125

Glu Lys Asn Leu Ser Gln Ile Gln  
 130 135

<210> 136  
 <211> 33  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 136

Met Tyr Trp Thr Thr Lys Leu Ile Ile Ser Ser Lys Lys Ile Gln Lys  
 1 5 10 15

Gln Gln Thr Lys Lys Lys Thr Arg Gly Lys Pro Gly Thr Lys Gly Ser  
 20 25 30

Arg

<210> 137  
 <211> 29  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 137

Met Met Thr Lys Thr Leu Leu Asn Glu Asn Ser Ile Val Cys Glu Thr  
 1 5 10 15

Leu Lys Lys Ser Leu Phe Ile Ser Phe Cys Arg Trp Asn  
 20 25

<210> 138  
 <211> 62  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 138

Met Gly Leu Pro Met Phe Ala Arg Leu Val Phe Glu Leu Leu Gly Ser  
 1 5 10 15

Lys Pro Ile Pro Thr His Leu Gly Pro Pro Gln Ser Ala Gly Asn Tyr  
 20 25 30

Arg His Glu Pro Leu His Leu Pro Ala Leu Val Thr Leu Asn Glu Leu  
 35 40 45

74

Leu Asn Leu Cys Ile Ser Ile Ser Leu Leu Ala Lys Trp Arg  
 50 55 60

<210> 139  
 <211> 84  
 <212> PRT  
 <213> Homo sapien

<400> 139

Met Ala Val Gly Arg Gly Leu Pro Gly Val Thr Ala Lys Leu Cys Val  
 1 5 10 15

His Arg Gln Ala Gly Arg Met Leu Gln Pro Cys Gly Val Gly Thr Val  
 20 25 30

Glu Ala Phe Leu Cys Val Ala Glu Asn Val Ser Gln Ile Ser Gly Asn  
 35 40 45

Trp Asp Arg Lys Val Pro Arg Gly Ala Cys Met Gly Arg Leu Gln Lys  
 50 55 60

Val Ser Pro His Phe Met Phe Val Ile Ala Ala Gln Asp Arg Gln Thr  
 65 70 75 80

Pro Arg Gly Trp

<210> 140  
 <211> 72  
 <212> PRT  
 <213> Homo sapien

<400> 140

Met Leu Ile Lys His Phe Thr Phe Ile Ile Lys Tyr Val Ala Met Phe  
 1 5 10 15

Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe  
 20 25 30

Phe Phe Phe Ser Leu Ser Pro Ser Phe Phe Phe Phe Tyr Ser Pro Ser  
 35 40 45

Gly Thr Pro Arg Gly Gly Glu Gly Asp Arg Gly Thr Arg Arg Glu Gly  
 50 55 60

Ala Arg Arg Glu Arg Ala Arg Arg  
 65 70

75

<210> 141  
 <211> 76  
 <212> PRT  
 <213> Homo sapien

<400> 141

Met Gly Lys Lys Ala Leu Asp Gln Leu Arg Ile Leu Arg Arg Leu Pro  
 1 5 10 15

Ser Gln Gly Trp Pro Val Lys Gly Cys Ile Leu His Thr Arg Ile Asp  
 20 25 30

Leu Thr Gln Gln Gln Arg Glu Lys Thr Ser Gln Ala Gln Ser Leu Ser  
 35 40 45

Pro Cys Gly Ser Ile Phe Thr Ile Ser Val Ser Cys Arg Gln Ser Asn  
 50 55 60

Trp Arg Tyr Gln Ala Ile Pro Gln Ile Leu Leu Phe  
 65 70 75

<210> 142  
 <211> 32  
 <212> PRT  
 <213> Homo sapien

<400> 142

Met Leu Ile Ser Arg Ile Ser Asn His Leu Leu Lys Phe Tyr Ala Leu  
 1 5 10 15

Ile Gly Val Ala Ile Gln Asp Phe Lys Lys Ile Phe Glu Pro Ser Gln  
 20 25 30

<210> 143  
 <211> 108  
 <212> PRT  
 <213> Homo sapien

<400> 143

Phe Leu Arg Gln Ser Leu Arg Ser Val Ala Gln Ala Gly Val Gln Ala  
 1 5 10 15

Arg His Leu Gly Ser Leu Gln Pro Leu Ser Leu Arg Phe Lys Ala Phe  
 20 25 30

Ser Cys Leu Ser Leu Leu Ser Ser Trp Asp Tyr Arg His Ala Pro Pro  
 35 40 45

His Pro Ala Asn Phe Phe Val Phe Leu Val Glu Met Gly Phe Thr Val  
 50 55 60

76

Leu Ala Arg Met Val Ser Ile Ser Ala Thr His Asp Pro Pro Ala Leu  
65 70 75 80

Ala Cys Gln Ser Ala Gly Ile Thr Gly Ala Arg Arg His Pro Arg Leu  
85 90 95

Ile His Ile His Phe Leu Ile Phe Glu Tyr Gln Ser  
100 105

<210> 144  
<211> 199  
<212> PRT  
<213> Homo sapien

<400> 144

Met Thr Thr His Glu Pro His Pro Arg His Lys His Ala Thr Thr Pro  
1 5 10 15

Ala Arg Thr His Pro Pro Asn His Glu Pro His Thr Pro Pro His Thr  
20 25 30

Thr Pro Thr Ser Pro Thr Thr Thr Pro Ala Thr Thr Pro Arg Thr His  
35 40 45

Thr Thr Thr Pro Thr Thr Ala Gln Thr Arg Arg Asp Arg Thr Ala Glu  
50 55 60

Lys Thr Thr Gln Arg Gly Gly Lys Glu Asp Asn Asp Ala Glu Gly Arg  
65 70 75 80

Arg Lys Arg Gly Pro Ile Thr Pro Pro Ala Ser Gly Ala Glu Ser Arg  
85 90 95

Gly Gly Leu Ala Arg Arg Ala Arg Trp Pro Pro Ala Asn Thr Thr Arg  
100 105 110

His Ala Thr Asn Asp Pro Thr His Gln Arg Thr Ala Gln Gln Gln Arg  
115 120 125

Arg Thr Ala Arg Asp Gln Arg Gly Thr Ala Asp Arg His Thr Asp Ala  
130 135 140

Arg Gly His Asp Gln Arg Arg Arg Thr Thr Gly Asp Asp Thr Arg Gln  
145 150 155 160

Ala Thr Gln Arg Ala Gln Pro Thr Gly Arg Glu Glu Lys Arg Gly Lys  
165 170 175

77

Lys Asn Ala Lys Ala Arg Pro Ala Ala Asn Arg Gly Ala Asn Gly Pro  
 180 185 190

Gln Ala Ala Ala Ala His Glu  
 195

<210> 145  
 <211> 88  
 <212> PRT  
 <213> Homo sapien

<400> 145

Met Arg Gly Ile Asn Pro Asp Pro Ser Val Cys Gly Ile Cys Asp Phe  
 1 5 10 15

Tyr Ser Ser Lys Val Ser Ile His Val Pro His Ser Glu Leu Ser Gln  
 20 25 30

Lys Asn Phe Ile Thr Leu Phe Ile Phe Phe Leu Arg Gly Lys Phe Lys  
 35 40 45

Gln Arg Lys Ser Leu Ala Gly Tyr Thr Gln Trp Leu Ile Gly Val Asp  
 50 55 60

Leu Arg Gly Gly Asp Asn Cys Val Tyr Ser Arg Ser His Thr Ser Pro  
 65 70 75 80

His Asn Tyr Tyr Arg Thr Asn Thr  
 85

<210> 146  
 <211> 63  
 <212> PRT  
 <213> Homo sapien

<400> 146

Met Trp Glu Gln Asn Phe Ile Cys Ala Phe Ile Val Glu Gln Glu Ser  
 1 5 10 15

His Leu Ala Leu Tyr Pro Ser Ser Leu Leu Tyr Asn Ser His Arg Asn  
 20 25 30

Val Ile Lys Leu Ala Ser Asn Trp Thr Arg Arg Lys Arg Trp Glu Thr  
 35 40 45

Pro Gly Ser Ile Ser Arg Val Arg Pro Pro Glu Lys Gly Ser Val  
 50 55 60

78

<210> 147  
 <211> 50  
 <212> PRT  
 <213> Homo sapien

<400> 147

Met Arg Pro Pro Ile Thr Leu Leu Gly Ala Arg Asp Lys Asn Lys Lys  
 1 5 10 15

Ser Trp Ala Val Pro Arg Gly Ala Ser Ala Trp Cys Pro Gly Gly Lys  
 20 25 30

Met Gly Asn Pro Ala His Asn Pro Pro Thr Thr Ile Pro Ala Gln Arg  
 35 40 45

Thr Arg  
 50

<210> 148  
 <211> 36  
 <212> PRT  
 <213> Homo sapien

<400> 148

Met Pro Gln Gly Lys Lys Tyr Asn Thr Tyr Ile His Lys Gln Lys Lys  
 1 5 10 15

Gln Glu Arg Ile Gln Met Ser Phe Asn Arg Gly Met Leu Thr Leu Met  
 20 25 30

Val Ala Tyr Ser  
 35

<210> 149  
 <211> 98  
 <212> PRT  
 <213> Homo sapien

<400> 149

Met Ser Ser Ser Ala Pro Thr Pro Trp Gly Ala Lys Gly Gly Glu Leu  
 1 5 10 15

Trp Arg Pro Glu Lys Pro Thr Phe Ser Thr His Gly Glu His Arg Tyr  
 20 25 30

Glu Pro His Trp Ser Asn Pro Gln Ala Leu Phe Phe Phe Leu Phe Phe  
 35 40 45

Phe Phe Phe Phe Phe Arg Lys Arg His Val Ile Tyr Phe Met Asn Ser

79

50

55

60

Ile Ser Arg Leu Ser Gly Asn Met Glu His Trp Gly Thr Asp Pro Ser  
65 70 75 80

Thr Glu Gly Phe Ala Ser Leu Leu Trp Phe Ser Cys Gln Leu Met Ile  
85 90 95

Arg Pro

&lt;210&gt; 150

&lt;211&gt; 94

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 150

Met Cys His Leu Leu Ile Phe Ile Arg Asn Leu Ser Leu Val Ala Thr  
1 5 10 15

Trp Pro Asn Thr Leu Gln Ser Met Ser Val Cys Leu Ser Val Cys Val  
20 25 30

Ser Leu Cys Val Cys Val Cys Val Cys Val Cys Val Cys Val Cys Val  
35 40 45

Cys Val Ser Pro His Ser Phe Ile Leu Ser Leu His Ser Ser Ile Ile  
50 55 60

Ile Asn Ile Arg Glu Ile His Arg Lys Tyr Ile Glu Lys Ile Thr Val  
65 70 75 80

Phe Ser Ile Lys Lys Lys Gln Leu Pro Ser Leu His Ser Phe  
85 90

&lt;210&gt; 151

&lt;211&gt; 260

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 151

Leu Arg Arg Ala Lys Ala His Glu Gly Leu Gly Phe Ser Ile Arg Gly  
1 5 10 15

Gly Ser Glu His Gly Val Gly Ile Tyr Val Ser Leu Val Glu Pro Gly  
20 25 30

Ser Leu Ala Glu Lys Glu Gly Leu Arg Val Gly Asp Gln Ile Leu Arg  
35 40 45

80

Val Asn Asp Lys Ser Leu Ala Arg Val Thr His Ala Glu Ala Val Lys  
50 55 60

Ala Leu Lys Gly Ser Lys Lys Leu Val Leu Ser Val Tyr Ser Ala Gly  
65 70 75 80

Arg Ile Pro Gly Gly Tyr Val Thr Asn His Ile Tyr Thr Trp Val Asp  
85 90 95

Pro Gln Gly Arg Ser Ile Ser Pro Pro Ser Gly Leu Pro Gln Pro His  
100 105 110

Gly Gly Ala Leu Arg Gln Gln Glu Gly Asp Arg Arg Ser Thr Leu His  
115 120 125

Leu Leu Gln Gly Gly Asp Glu Lys Lys Val Asn Leu Val Leu Gly Asp  
130 135 140

Gly Arg Ser Leu Gly Leu Thr Ile Arg Gly Gly Ala Glu Tyr Gly Leu  
145 150 155 160

Gly Ile Tyr Ile Thr Gly Val Asp Pro Gly Ser Glu Ala Glu Gly Ser  
165 170 175

Gly Leu Lys Val Gly Asp Gln Ile Leu Glu Val Asn Gly Arg Ser Phe  
180 185 190

Leu Asn Ile Leu His Asp Glu Ala Val Arg Leu Leu Lys Ser Ser Arg  
195 200 205

His Leu Ile Leu Thr Val Lys Asp Val Gly Arg Leu Pro His Ala Arg  
210 215 220

Thr Thr Val Asp Glu Thr Lys Trp Ile Ala Ser Ser Arg Ile Arg Glu  
225 230 235 240

Thr Met Ala Asn Ser Ala Gly Ser Gly His Ser Ala Arg Ser Asn Leu  
245 250 255

Gln Thr Pro Gly  
260

<210> 152  
<211> 95  
<212> PRT  
<213> Homo sapien



81

&lt;400&gt; 152

Met Trp Val Leu Val Leu Gly Ala Leu Leu Ala Gly Ile Ile Pro Leu  
 1 5 10 15

Cys Tyr Ser Pro Gly Ile Gln Arg Phe Leu Pro Pro Trp Gly Leu Pro  
 20 25 30

Pro Thr Ala Phe Cys Arg Gln Cys Val Phe Ala Leu Val Ser Cys Gly  
 35 40 45

Ala Arg Gly Ser Arg Ser Ala Gly Gly Val Ser Gly Gly Ala Pro Arg  
 50 55 60

Cys Ala Pro Leu Phe Ile Trp Gly Ile Cys Val Cys Gly Gly Ser Pro  
 65 70 75 80

Pro Trp Phe Ala Val Cys Arg Ala Cys Gly Ser Pro Arg Ser Val  
 85 90 95

&lt;210&gt; 153

&lt;211&gt; 62

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 153

Met Phe Ser Val Val Val Trp Cys Leu Leu Val Arg Cys Val Val Val  
 1 5 10 15

Asn Cys Gly Glu Leu Trp Arg Gly Ile Thr Asn Val His Pro Gly Gly  
 20 25 30

Pro Ala Tyr Glu Pro Glu Ala Thr Pro Gln Ala Phe Phe Phe Cys Phe  
 35 40 45

Phe Phe Leu Leu Val Lys Glu Pro Ser Phe Ile Ile Lys Gln  
 50 55 60

&lt;210&gt; 154

&lt;211&gt; 65

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 154

Met Arg Leu Ile Gln Lys Arg Arg Ile Tyr Pro Ser Arg Lys Thr Glu  
 1 5 10 15

Ile Asn Ser Ser Ser Pro Phe Thr Tyr Pro Pro Tyr Thr His Thr Tyr  
 20 25 30

82

Asn Thr His Thr His Thr His Thr Glu Arg Glu Arg Asp Leu Pro Gly  
 35 40 45

Gly Ile His His Leu Arg Arg Ser Ser Asn Ala Ile Asn Gly Pro Phe  
 50 55 60

Ala  
 65

<210> 155  
 <211> 51  
 <212> PRT  
 <213> Homo sapien

<400> 155

Met Ile Cys Ile Pro Leu Arg Lys Asn Ser Ser Trp Glu Phe Ile Arg  
 1 5 10 15

Leu Phe Phe Ile Pro Ala His Lys Lys Lys Leu Leu Ala Leu Leu  
 20 25 30

Leu Lys Thr Glu Glu Pro Gln Glu Lys Ile Ser Phe Ser Tyr Arg Ala  
 35 40 45

Lys Ile Lys  
 50

<210> 156  
 <211> 129  
 <212> PRT  
 <213> Homo sapien

<400> 156

Met Leu Leu Glu Arg Pro Gln Cys Asp Gly Cys Ala Arg Ala Gly Thr  
 1 5 10 15

Ala Phe Phe Phe Phe Phe Phe Leu Gly Asn Gly Ile Leu Leu Cys His  
 20 25 30

Pro Gly Trp Ile Lys Val Ala Gln Pro Trp Phe Thr Glu Thr Ser Ala  
 35 40 45

Ser Trp Val Val Phe Lys Asn Ile Leu Leu Phe Ser Cys Val Leu Ser  
 50 55 60

Ala Ser Pro Lys Leu Ala Val Gly Leu Thr Gly Leu Ala Thr Thr Ala  
 65 70 75 80

83

Thr Gln Leu Asn Phe Val His Val Phe Ser Lys Ala Arg Gly Phe Ser  
85 90 95

Leu Asn Leu Phe Gly Pro Gly Val Val Ser Arg Leu Leu Arg Glu Pro  
100 105 110

Gln Val Thr Pro Ser Val Pro Ser Arg Leu Leu Lys Met Trp Leu Val  
115 120 125

Tyr

<210> 157  
<211> 71  
<212> PRT  
<213> Homo sapien

&lt;400&gt; 157

Met Ile Arg Gln Ala Val Phe Asn Ala Val Tyr Asn Cys Phe Ile Ile  
1 5 10 15

Ser Cys Ser Asp Cys Ser Leu Leu Val Cys Arg Asn Thr His Leu Phe  
20 25 30

Cys Asp Pro Cys Leu Gln Pro His Ser Leu Ile Ile Phe Ile Leu Ile  
35 40 45

Ala Ile Leu Arg Met Cys Ser Ile Tyr Arg Asp Pro Ile Ile Leu Val  
50 55 60

Glu Leu Lys Ile Cys Leu Cys  
65 70

<210> 158  
<211> 69  
<212> PRT  
<213> Homo sapien

&lt;400&gt; 158

Met Arg Leu Pro Leu His His Val Leu Pro Leu Arg Asp Leu Ser Phe  
1 5 10 15

Gln His Tyr Ser Cys Lys Leu Gln Trp His Ser Thr Thr Phe Ile Pro  
20 25 30

Ser Ser Cys His Ser Leu Phe Phe His Ser Phe Leu Thr Val Cys Thr  
35 40 45

Pro Met Tyr Ala Ala Ile Phe Ile Ile Leu His Phe Leu Tyr Leu Ser

84

50

55

60

Ile Pro Asn Ile Leu  
65

<210> 159  
<211> 57  
<212> PRT  
<213> Homo sapien

<400> 159

Met Ser His Cys Thr Gln Pro Gly Glu Ser Phe Ile Met Gly Tyr Glu  
1 5 10 15

Val Tyr Arg Leu His Ser Asp Ser Thr Lys Leu Asp Phe Met Arg Ile  
20 25 30

Gln Leu Gln Leu Thr Phe Thr Ser Gly Leu Thr Leu Lys Arg Lys Ile  
35 40 45

Val Ser Gln Lys Asp Leu Trp Tyr Met  
50 55

<210> 160  
<211> 102  
<212> PRT  
<213> Homo sapien

<400> 160

Met Tyr His Phe Ser Thr Leu Arg Ala Cys Leu Gly Pro Phe Phe Cys  
1 5 10 15

Val Arg Cys Leu Gln Thr Ile Leu Thr Ile Leu Glu Arg Ala Leu Pro  
20 25 30

Arg Arg Glu Ser Arg Gly Thr Phe Leu Phe Ser Gln Lys Lys Pro Arg  
35 40 45

Val Ile Arg Phe Pro Pro Pro Gly Gly Gly Leu Leu Asn Gln Glu Val  
50 55 60

Asp Leu Leu Ala Ser Ile Ser Val Tyr Asn Pro Gln Pro Ser Gly Val  
65 70 75 80

Thr Thr Gly Leu Gln Arg Val Cys Asp Asn Val Ser Asn Ala Glu Lys  
85 90 95

Lys Thr Pro Ser Pro Val  
100

85

<210> 161  
 <211> 70  
 <212> PRT  
 <213> Homo sapien

<400> 161

Met Val Met Cys Gln Pro Glu Gly Asn Val Tyr Ala Val Leu Arg Ser  
 1 5 10 15

Pro Leu Phe Leu Glu Asn Gln Gln Asn Arg Ala Asp His Leu Ala Tyr  
 20 25 30

His Phe Cys Val Leu Leu Val Pro Gly Ile Gly Leu Trp Phe Asp His  
 35 40 45

Cys Cys Asp His Cys Ser Ala Asp Cys Asp Leu Gln Asn Thr Glu Ser  
 50 55 60

Lys Leu Gln Ser Pro Trp  
 65 70

<210> 162  
 <211> 59  
 <212> PRT  
 <213> Homo sapien

<400> 162

Met Gly Cys His Lys Ser Gly Thr Gly Gly Phe Leu Ser Arg Gly Lys  
 1 5 10 15

Arg Thr Glu Pro Ala His His Val Met Pro Cys His Leu Arg Ile Leu  
 20 25 30

His Ser Ser His Gln Glu Glu Gly Pro His Gln Met Gln Pro Leu Asn  
 35 40 45

Phe Glu Leu Leu Ser Leu Gln Ser Cys Gln Lys  
 50 55

<210> 163  
 <211> 84  
 <212> PRT  
 <213> Homo sapien

<400> 163

Met Thr Thr Gln Thr Gly Asn Gln Leu Asp Ala His Gly Gly Ser Ala  
 1 5 10 15

86

Gln Ala Leu Phe Cys Phe Phe Leu Phe Phe Phe Tyr Leu Lys Tyr Leu  
                   20                                  25                                  30

Val Leu Asn Leu Val Gln Leu Asn His Trp Glu Phe Glu Phe Leu Phe  
                   35                                  40                                  45

Lys Ser Cys Leu Trp Ser Ala Ser Tyr Gly Lys Pro Leu His Trp Ile  
                   50                                  55                                  60

Pro Ser Thr Lys Thr Arg Leu Leu Lys Phe Lys Cys Gln Trp Gly Arg  
                   65                                  70                                  75                                  80

Trp Glu Ala Ala

<210> 164  
 <211> 41  
 <212> PRT  
 <213> Homo sapien

<400> 164

Met Cys His His His Gly Asn His Ala Phe Trp Ala Pro Leu Gly Val  
                   1                                  5                                  10                                  15

Thr Ala Pro Ser Ala Val Leu Phe Cys Phe Val Phe Leu Phe Cys Phe  
                   20                                  25                                  30

Phe Ser Gln Leu Gly Lys Phe Asn Ile  
                   35                                  40

<210> 165  
 <211> 51  
 <212> PRT  
 <213> Homo sapien

<400> 165

Met Arg Leu Phe Phe Thr Ser Leu Ser Gln Gly Cys Phe Phe Leu Val  
                   1                                  5                                  10                                  15

Ile Cys Leu Leu Cys Phe Ile Arg Tyr Phe Ala Gln Ile Lys His Ser  
                   20                                  25                                  30

Pro Gly Ala Gln Lys Lys Lys Lys Lys Lys Lys Lys Arg Pro Arg  
                   35                                  40                                  45

Arg Asp His  
                   50

<210> 166

87

<211> 31  
 <212> PRT  
 <213> Homo sapien

<400> 166

Met Trp Leu Val Phe Pro Leu Tyr Ile Lys Met Leu Leu Ser Gly Ile  
 1 5 10 15

Ala Gln Asp Pro Gln Thr Asn Arg Asp Tyr Leu Pro Arg Thr Lys  
 20 25 30

<210> 167  
 <211> 74  
 <212> PRT  
 <213> Homo sapien

<400> 167

Met Ser His Thr Pro Val Thr Tyr Pro Ala Arg Gly Ser Gly Asn Ser  
 1 5 10 15

Pro Ile Ser Ala Cys Val Ile Phe Gln Trp Trp Cys Ser Glu Val Cys  
 20 25 30

Leu Pro Met Ala Ser Gln Pro Val Ala Gly Val Leu Trp Met Gly Leu  
 35 40 45

Pro Ser Met Val Pro Leu Leu Ser Gln Glu Thr Gly Glu Asn Glu Ala  
 50 55 60

Phe Ser Arg Val Phe Glu Val Ala Asn Ala  
 65 70

<210> 168  
 <211> 229  
 <212> PRT  
 <213> Homo sapien

<400> 168

Met Ser Leu Leu Cys Leu Leu Leu Ser Phe Leu Leu Phe Tyr Phe Ser  
 1 5 10 15

Ala Leu Val Phe Ser Tyr Ala Ser Leu Phe Pro Leu Val Ala Ser Cys  
 20 25 30

Cys Ser Val Leu Phe Val Phe Met Arg Ser Gly Gly Leu Cys His Val  
 35 40 45

Cys Gly Leu Ala Leu Phe Val Cys Phe Leu Leu Val Gly Leu Leu Arg  
 50 55 60

88

Leu Arg Ser Pro Leu Tyr Thr Pro Leu Ser Val Ala Phe Arg His Ser  
65 70 75 80

Arg Arg Val Ser Phe Cys Cys Ala Phe Arg Val Ser Val Val Ser  
85 90 95

Leu Arg His Val Val Cys Val Arg Cys Val Ser Phe Met Val Leu Phe  
100 105 110

Ser Phe Ser Ser Leu Phe Ala Val Leu Leu Phe Val Arg Ser Phe Ser  
115 120 125

Leu Trp Phe Ala Phe Cys Ser Leu Val Pro Phe Leu Cys Ala Leu Val  
130 135 140

His Val Leu Phe Phe Arg Leu Leu Phe Leu Ser Ser Phe Val Val Leu  
145 150 155 160

Leu Ile Met Leu Phe Phe Val Leu Leu Phe Leu Thr Leu Leu Ser Cys  
165 170 175

Phe Ser Leu Ser Arg Pro Phe Cys Ser Phe Leu Cys Leu Tyr Ala Ser  
180 185 190

Met Ser Val Cys Leu Gly Arg Ala Arg Gly Cys Val Ile Ala Gly Ser  
195 200 205

Gly Arg Leu Leu Ala Ile Tyr Arg Leu Met Arg Cys Leu Val Ser Pro  
210 215 220

Cys Leu Leu Leu Ala  
225

<210> 169  
<211> 34  
<212> PRT  
<213> Homo sapien

<400> 169

Met Leu Gly Phe Leu Ala His Phe Gln Arg Phe Ala Arg Lys Lys Val  
1 5 10 15

Pro Lys His Gln Leu Ile Ser Ser Ser Leu His Val Gly His Gly Asn  
20 25 30

Ile Ser



89

<210> 170  
 <211> 51  
 <212> PRT  
 <213> Homo sapien

<400> 170

Met Gly Met Gly Ala Gly Lys Pro Phe His Thr Arg Thr Ser Cys Arg  
 1 5 10 15

Pro Trp Leu Pro Pro His Leu Phe Phe Phe Phe Phe Ser Glu Val  
 20 25 30

Asn Leu Asp Leu Cys Leu Phe Thr Pro His Tyr Val Lys Thr Gly Ala  
 35 40 45

Ser Phe Leu  
 50

<210> 171  
 <211> 46  
 <212> PRT  
 <213> Homo sapien

<400> 171

Met Cys Pro Cys Lys Arg Val Phe Ala Asp Thr Thr Ser Phe Ile Thr  
 1 5 10 15

Gln Gly Pro Gln Phe Ile Pro Phe Pro Gln Glu Val Pro Pro Pro Leu  
 20 25 30

Ser Glu Gly Lys Asn Phe Pro Ala Val Asn Tyr Arg Ala Tyr  
 35 40 45

<210> 172  
 <211> 45  
 <212> PRT  
 <213> Homo sapien

<400> 172

Met Ala Val Ala Phe Gln Ser Leu Ile Pro Trp Gly Leu Gln Leu Cys  
 1 5 10 15

Val Asn Lys Val Ala Ala Asp Glu Leu Val Leu Thr Arg Lys Met Lys  
 20 25 30

Ala Lys Tyr Ala Ser Ile Ser Ser Arg Gln His Thr Asp  
 35 40 45

<210> 173

90

<211> 59  
 <212> PRT  
 <213> Homo sapien

<400> 173

Met Met Lys Leu Arg Trp Arg Ile Leu Lys Pro Gly Ala Glu Val Thr  
 1 5 10 15

Met Lys Arg Asn Val Gln Leu His Ser Ser Leu Gly Thr Glu Glu Asp  
 20 25 30

Leu His Arg Lys Lys Lys Lys Lys Lys Ser Leu Val His Gly Ile  
 35 40 45

Cys Pro Cys Val Asn Val Ser Arg Gln Ser Gln  
 50 55

<210> 174  
 <211> 59  
 <212> PRT  
 <213> Homo sapien

<400> 174

Met Lys Ile Gly Pro Met Phe Thr Trp Val Glu Thr Tyr Ile Thr His  
 1 5 10 15

Leu Gln Leu Gly Pro Leu Cys Gln Thr Ser Phe Gln Thr Gln Arg His  
 20 25 30

Ala Gly Ala Ser Ser Leu Ser Ile Asn Gly Ser Ala Val Gly Met Ser  
 35 40 45

Ala Val Gly Gly Leu Leu Leu Gly Glu Ser His  
 50 55

<210> 175  
 <211> 74  
 <212> PRT  
 <213> Homo sapien

<400> 175

Met Phe Thr Ile His Arg Val Arg Ile Pro His Lys Ile Phe Arg Arg  
 1 5 10 15

Pro His Ile Leu Ile Gly Ser Val Pro Ile Pro Ser Leu Phe Arg Gly  
 20 25 30

Pro Lys Leu Phe Phe Thr Ser Ser Ser Ala Ile Met Gly Asn Pro Phe  
 35 40 45

91

Val Val Tyr Thr His Lys Arg Val Gly Arg Trp Asn Lys Pro Leu Tyr  
 50 55 60

Val Met Leu Leu Met Lys Val Ile Ser Leu  
 65 70

<210> 176  
 <211> 73  
 <212> PRT  
 <213> Homo sapien

<400> 176

Met Gln Ser Gln Leu His Ser Tyr Phe Phe Glu Arg Arg Ala Arg Phe  
 1 5 10 15

His Thr Leu Cys Ala Arg Asn Ile Asn Ile Ser Ser Ser Leu Gln Glu  
 20 25 30

Glu Val Pro Thr Ile Leu Val Met Pro His Ser Lys Lys Thr Ile Phe  
 35 40 45

Val Glu Lys Leu Phe Phe Gly Ala Thr Ala Phe Ala Leu Lys Asn Cys  
 50 55 60

Cys Leu Phe Thr Pro Pro Thr Tyr Phe  
 65 70

<210> 177  
 <211> 129  
 <212> PRT  
 <213> Homo sapien

<400> 177

Met Ala Val Ser Val Ser Leu Cys Ser Ser Pro Arg Cys Leu Ser Leu  
 1 5 10 15

Leu Phe Val Ala Ser Ala Arg Ala Thr Arg Pro Leu Leu Val Leu Ser  
 20 25 30

Val Val His Ser Arg Ser Trp Leu Val Leu Ser Cys Ala Phe Leu Ser  
 35 40 45

Ser Gly Ser Cys Pro Arg Arg Leu Leu Val Ser Cys Tyr Arg Val Gly  
 50 55 60

Cys Val Ser Pro Ser Gly Ala Ser Phe Ser Ser Ser Ala Ser Ser Ser  
 65 70 75 80

92

Ala Pro Phe Cys Trp Val Gly His Phe Cys Pro Arg Gly Asp Ser Arg  
85 90 95

Val Ile Pro Gly Glu Ser Thr Met Gly Met Arg His Thr Thr Cys Tyr  
100 105 110

Arg Arg Thr His Gly Arg Trp Phe Val Gly Cys Phe Val Val Val Cys  
115 120 125

Phe

<210> 178  
<211> 52  
<212> PRT  
<213> Homo sapien

&lt;400&gt; 178

Met Leu Gly Ile Val Gly Pro Gly Thr His Phe Thr Pro Gly Asp Tyr  
1 5 10 15

Arg Phe Gly Ala Leu Gly Val Ala Pro Ser Arg Phe Arg Cys Val Tyr  
20 25 30

Glu Cys Val Ser Ser Lys Arg Lys Lys Gly Thr Leu Asn Asn Pro Leu  
35 40 45

Gly His Ser Gly  
50

<210> 179  
<211> 90  
<212> PRT  
<213> Homo sapien

&lt;400&gt; 179

Met Met Phe Tyr Thr Gln Thr Pro Val Phe Val Pro Phe Val Pro Pro  
1 5 10 15

Asn Asn Ile Cys Pro Leu Ile Met Asn Tyr Tyr Thr Gln Ser Ala Ile  
20 25 30

Pro Gly Val Tyr Thr Pro Tyr Leu Arg Tyr Lys Phe Ser Pro Lys Ile  
35 40 45

Val Lys Lys Lys Lys Pro Pro Phe Leu Asn Asn Lys Thr Phe Val Pro  
50 55 60

Trp Asn Lys Arg Lys Phe Leu Pro Leu Pro Lys Lys Lys Lys Lys

|                   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| 93                |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| 65                |     |     |     | 70  |     |     |     | 75  |     |     |     | 80  |     |     |     |  |
| Lys               | Lys | Gly | Gly | Gly | Thr | Cys | Pro | Ala | Ala |     |     |     |     |     |     |  |
|                   |     |     |     | 85  |     |     |     | 90  |     |     |     |     |     |     |     |  |
| <210> 180         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <211> 142         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <212> PRT         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <213> Homo sapien |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <400> 180         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Met               | Ser | Met | Ser | Cys | Gly | Ala | Gly | Ala | Pro | Leu | Arg | Val | Cys | Val | Ser |  |
| 1                 |     |     |     | 5   |     |     |     |     | 10  |     |     |     | 15  |     |     |  |
| Trp               | Trp | Leu | Trp | Val | Gly | Gly | Arg | Val | Gly | Ala | Val | Val | Arg | Pro | Arg |  |
|                   |     |     | 20  |     |     |     |     | 25  |     |     |     | 30  |     |     |     |  |
| Ala               | Leu | Trp | Ser | Ala | Trp | Gly | Ala | Val | Gly | Gly | Gly | Leu | Leu | Cys | Val |  |
|                   |     | 35  |     |     |     |     | 40  |     |     |     | 45  |     |     |     |     |  |
| Val               | Ala | Leu | Phe | Trp | Leu | Cys | Ala | Gly | Arg | Arg | Gly | Ala | Arg | Leu | Pro |  |
|                   |     | 50  |     |     |     |     | 55  |     |     |     | 60  |     |     |     |     |  |
| Pro               | Ser | Pro | Cys | Gly | Ala | Val | Ala | Val | Ala | Ala | Val | Asp | Ala | Gly | Ala |  |
| 65                |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |  |
| Ala               | Gly | Gly | Val | Val | Arg | Gly | Gly | Gly | Val | Val | Val | Val | Gly | Arg | Trp |  |
|                   |     |     |     | 85  |     |     |     |     | 90  |     |     |     | 95  |     |     |  |
| Leu               | Gly | Arg | Leu | Gly | Trp | Val | Val | Gly | Arg | Val | Cys | Ala | Arg | Gly | Pro |  |
|                   |     |     | 100 |     |     |     |     | 105 |     |     |     | 110 |     |     |     |  |
| Cys               | Leu | Cys | Arg | Gly | Gly | Ala | Trp | Ala | Gly | Ala | Ala | Gly | Arg | Gly | Gly |  |
|                   |     | 115 |     |     |     |     | 120 |     |     |     | 125 |     |     |     |     |  |
| Gly               | Gly | Arg | Arg | Gly | Arg | Arg | Gly | Arg | Ala | Arg | Gly | Pro | Gly |     |     |  |
|                   |     | 130 |     |     |     |     | 135 |     |     |     | 140 |     |     |     |     |  |
| <210> 181         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <211> 80          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <212> PRT         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <213> Homo sapien |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| <400> 181         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Met               | Ser | Arg | Arg | Gly | Pro | Pro | Pro | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe |  |
| 1                 |     |     |     | 5   |     |     |     |     | 10  |     |     |     | 15  |     |     |  |
| Phe               | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe | Phe |  |
|                   |     |     | 20  |     |     |     |     | 25  |     |     |     | 30  |     |     |     |  |

94

Phe Phe Phe Phe Phe Lys Lys Lys Lys Lys Leu Leu Phe Ile Lys Lys  
 35 40 45

Gly Gly Gly Gly Ala Arg Gly Gly Gly Gly Arg Ala Pro Gly Gly Gly  
 50 55 60

Gly Gly Gly Glu Lys Thr Thr Lys Lys Arg Arg Thr Thr Ser Gly Pro  
 65 70 75 80

<210> 182

<211> 72

<212> PRT

<213> Homo sapien

<400> 182

Met Leu Glu Arg Arg Ser Val Met Asp Glu Arg Arg Pro Gly Arg Phe  
 1 5 10 15

Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Leu Glu  
 20 25 30

Lys Lys Phe Phe Lys Asn Pro Gln Lys Phe Pro Gly Gln Gly Gly Leu  
 35 40 45

Pro Pro Gly Lys Lys Lys Lys Lys Lys Lys Ile Trp Ala Leu Trp Gly  
 50 55 60

Leu Pro Leu Ser Leu Val Gly Gly  
 65 70

<210> 183

<211> 95

<212> PRT

<213> Homo sapien

<400> 183

Met Arg Pro Pro Lys Phe Tyr Ser Leu Leu Asn Val Ser Pro His Ser  
 1 5 10 15

Arg Ala Leu Ser Ile Ala Pro Ser Thr Lys Lys Thr Ser Asn Arg Gly  
 20 25 30

Glu Asp Val Arg Arg Gly Glu Val Pro Pro Arg Ala His Ser Arg Cys  
 35 40 45

Lys His Cys Thr Thr Thr Pro His Pro Phe Gly Leu Cys Thr Thr Phe  
 50 55 60

95

Ser Thr Gly Gly Thr Thr Thr Phe Cys Arg Ser Ser Gln Thr Leu Ser  
65 70 75 80

Cys Leu Pro Ser Thr Pro Leu Leu Leu Pro Trp Val Leu Leu Cys  
85 90 95

<210> 184  
<211> 17  
<212> PRT  
<213> Homo sapien

<400> 184

Met Gly Glu Asp Lys Gln Asp Leu Phe Ala Phe Ala Ala Leu Ile Phe  
1 5 10 15

Leu

<210> 185  
<211> 71  
<212> PRT  
<213> Homo sapien

<400> 185

Met Ala Ala Asp Pro Ala Ser Ala Gln Gly Asp Ser Gly Thr Gly Tyr  
1 5 10 15

Val Ser Cys Leu Leu Ser Ile Phe Ala Gly Cys Ala Leu Gln Trp Cys  
20 25 30

Ala Leu Leu Leu Leu Cys Leu Phe Phe Leu Arg Leu Phe Phe Gly  
35 40 45

Ile Leu Trp Arg Val Thr Pro Val Pro Thr Gly Thr Pro Phe Ala Pro  
50 55 60

Glu Ile Met Pro Pro Thr Phe  
65 70

<210> 186  
<211> 59  
<212> PRT  
<213> Homo sapien

<400> 186

Met Ala Leu Ser Leu Ala Ala Trp Thr Leu Leu Glu Glu Cys Val Ser  
1 5 10 15

Ser Arg Cys Leu Pro Thr Val Met Gly Gly Ser Leu Phe Ile Gly Leu

96

20

25

30

Leu Leu Cys Leu Leu Ala Ser Met Phe Gly His Val Val Ser Pro Ser  
 35 40 45

Trp Phe His Thr Tyr Trp Asn Leu Val Tyr Pro  
 50 55

&lt;210&gt; 187

&lt;211&gt; 80

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 187

Pro Arg Lys Ala Leu Phe Thr Tyr Pro Lys Gly Ala Ala Glu Met Leu  
 1 5 10 15

Glu Asp Gly Ser Glu Arg Phe Leu Cys Glu Ser Val Phe Ser Tyr Gln  
 20 25 30

Val Ala Ser Thr Leu Lys Ala Val Lys His Asp Gln Gln Val Ala Arg  
 35 40 45

Met Glu Lys Leu Ala Gly Leu Val Glu Glu Leu Glu Ala Asp Glu Trp  
 50 55 60

Arg Phe Lys Pro Ile Glu Gln Leu Leu Gly Phe Thr Pro Ser Ser Gly  
 65 70 75 80

&lt;210&gt; 188

&lt;211&gt; 105

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 188

Met Arg Thr Met Met Thr Cys Asp Lys Ile His His Val Ser Ile Ser  
 1 5 10 15

Gln Ser Leu Gln Ile Gln Ser His Asn Glu Pro Leu Met Gln Gln Ser  
 20 25 30

His Pro His Ser Leu Ile Ser Leu Gly Asn Ile Thr Ala Tyr Thr Met  
 35 40 45

Asn Asn Pro Leu Arg Tyr Ala Asp Ser Ser His His Ser Val Glu Asn  
 50 55 60

Ser Ile Leu Leu Thr Val Arg Pro Thr Val Leu Phe Pro Arg Ala Ser  
 65 70 75 80



97

Val Glu Leu Gln Asn Arg Pro Ser Cys Asp Gln Pro Ser Gln Arg Leu  
                             85                            90                            95

Met Ser Gln Phe Val Ala Leu Asp Ser  
                             100                            105

<210> 189  
 <211> 83  
 <212> PRT  
 <213> Homo sapien

<400> 189

Met Cys Glu Ser Leu Ala Phe Leu Leu Leu Gln Phe Gly Tyr Phe Ala  
 1                            5                            10                            15

Leu Ile Ser Phe Val Asn Ser Ile Leu Tyr Ser Phe Asp Arg Arg Ala  
                             20                            25                            30

Tyr Cys Asn Lys Val Lys Ile Ile Ala Gln Lys Ile Leu His Ile Phe  
                             35                            40                            45

Ser Thr Asn Pro Tyr Cys Phe Leu Pro Thr Lys Asp Leu Tyr Tyr Ser  
                             50                            55                            60

Lys Cys Val Ser Thr Cys Leu Ala Leu Tyr Pro Gln Arg Lys Lys Cys  
 65                            70                            75                            80

His. Leu Leu

<210> 190  
 <211> 40  
 <212> PRT  
 <213> Homo sapien

<400> 190

Met Ile Thr Pro Leu His Ser Ser Leu Gly Lys Ser Asp Thr Gln Pro  
 1                            5                            10                            15

Lys Lys Asn Asn Lys Lys Lys Lys Lys Lys Asn Thr Trp Gly Ile Pro  
                             20                            25                            30

Trp Gly Lys Gly Cys Ser Gly Val  
                             35                            40

<210> 191  
 <211> 75  
 <212> PRT

98

&lt;213&gt; Homo sapien

&lt;400&gt; 191

Met Thr Asn Asn Thr Pro Lys Phe Phe Phe Phe Phe Phe Phe Leu  
 1 5 10 15

Gly Glu Thr Glu Ser Leu Thr Leu Ser Pro Arg Leu Glu Cys Ser Gly  
 20 25 30

Glu Ile Ser Ala His Cys Asn Leu Arg Leu Leu Asp Ser Cys Asp Ser  
 35 40 45

Pro Val Ser Ser Phe Pro Ser Ser Trp Gly Tyr Arg Arg Gly Pro His  
 50 55 60

Leu Pro Gly Asp Pro Ser His Cys Ala Val Arg  
 65 70 75

&lt;210&gt; 192

&lt;211&gt; 67

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 192

Met His Phe Cys Gln Leu Leu Arg Thr Ser Ser Leu Ile Gly Met Cys  
 1 5 10 15

Trp Val Leu Arg Phe Ser Tyr Phe Phe Lys Leu Cys Leu Glu Phe Lys  
 20 25 30

Asn Tyr Thr Ser Leu Asn Tyr Met Pro Asn Ser Trp Pro Thr Gln Met  
 35 40 45

Lys Val Leu Val Leu Leu Ser Val Ile Pro Gly Leu Cys Gly Asn Leu  
 50 55 60

Asn Thr Ser  
 65

&lt;210&gt; 193

&lt;211&gt; 47

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 193

Met Trp Thr Gly Asn Asn Gln Ile Val His Pro Thr Gly Thr Thr Leu  
 1 5 10 15

Trp Pro Thr Glu Leu Pro Ala Arg Leu Phe Phe Val Phe Phe Cys Phe

99

20

25

30

Phe Leu Ile Lys Cys Leu Tyr Phe Ile Lys Lys Thr Ser Pro Phe  
 35 40 45

&lt;210&gt; 194

&lt;211&gt; 68

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 194

Met Ala His Gly Val Pro Leu Ala Leu Pro Val Val Pro Ala Trp Trp  
 1 5 10 15

Gly Cys Ser Arg Arg Leu Leu Ala Pro Gly Phe Ala Thr Pro Leu Leu  
 20 25 30

Arg Gly Phe Ala Pro Leu Leu His His Arg Arg Gly Arg Lys Asn Glu  
 35 40 45

Lys Lys Glu Glu Phe Leu Arg Val Thr Met Met Asn Thr Trp Gly Leu  
 50 55 60

Ala Leu Leu Val  
 65

&lt;210&gt; 195

&lt;211&gt; 68

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 195

Met Thr Asn His Asp Thr Thr Val Gly Val Leu Ile Tyr His Thr His  
 1 5 10 15

His Lys Leu Leu Thr Thr Ile Ile Asn Ile Ser Leu Phe Phe Ser Gly  
 20 25 30

Glu His Asn Asn Thr Thr Leu Phe Phe Glu Thr His Thr Leu Phe Thr  
 35 40 45

Thr Thr Phe Phe Phe Phe His Ser Pro Ser Pro Pro His Phe Pro Gly  
 50 55 60

Phe Phe Phe Leu  
 65

&lt;210&gt; 196

&lt;211&gt; 122

100

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 196

Met Asp Ala Ala Arg Ala Gly Lys Lys Lys Lys Lys Lys Lys Lys  
 1 5 10 15

Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys  
 20 25 30

Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Gly Gly Gly Phe Val  
 35 40 45

Pro Ser Ser Pro Leu Phe Leu Phe Ser Ile Thr Thr Phe Pro Arg Asp  
 50 55 60

Arg Ala Ala Arg Gly Gly Asp Thr Leu Tyr Tyr Ile Glu Glu Gly Asp  
 65 70 75 80

Arg Arg Tyr Ser Ser Lys Arg Ala Glu Asn Ile Ala Lys Ile Gly Trp  
 85 90 95

Leu Pro Gly Glu Thr Ile Glu Val Val Ala Thr Ile Leu Glu Pro Phe  
 100 105 110

Ala Cys Arg Leu Val His Thr Thr Pro Gln  
 115 120

&lt;210&gt; 197

&lt;211&gt; 84

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 197

Met Cys Leu Leu Ala Pro Cys Pro Glu Thr Pro Glu Ser Ser Trp Val  
 1 5 10 15

Val Lys Glu Ile Pro Trp Ser Ser Gln Val Pro Gly Ala Thr Cys Trp  
 20 25 30

Gly Phe Pro Gly His Arg Leu Ser Leu Lys Ala Cys Arg His Cys Ala  
 35 40 45

Thr Val Val Pro Val Arg Pro Ser Trp Gly His Gly Glu Arg Asp Ile  
 50 55 60

Ala Ile Pro Glu Ile Pro Gln Ser Val Met Cys Asp Leu Arg Ile Leu  
 65 70 75 80

101

Leu Arg Thr Pro

<210> 198  
 <211> 84  
 <212> PRT  
 <213> Homo sapien  
 <400> 198

Met Asn Lys Leu His Trp Gln Trp Pro Leu Ser Ser Arg Arg Arg Gln  
 1 5 10 15

Leu Met Asp Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe  
 20 25 30

Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Leu  
 35 40 45

Gly Gly Gly Thr Gly Glu Gln Gly Gly Arg Ala Gly Gly Glu Cys Val  
 50 55 60

Leu Pro Pro Pro Pro Pro Gln Lys Lys Lys Lys Lys Asn Ser Ile Asn  
 65 70 75 80

Lys Lys Lys Lys

<210> 199  
 <211> 134  
 <212> PRT  
 <213> Homo sapien  
 <400> 199

Met Pro Leu His Ser Ser Leu Gly Asn Arg Val Arg Pro Cys Pro Ser  
 1 5 10 15

Thr Leu Gly Gly Arg Gly Ala Gln Leu Glu Ile Ser Leu Gly Asn Ile  
 20 25 30

Val Lys Leu Asp Leu Tyr Lys Lys Lys Lys Lys Lys Lys Ser Arg Val  
 35 40 45

Trp Trp Cys Ala Pro Val Val Pro Ala Thr Gly Lys Leu Arg Trp Glu  
 50 55 60

Asp His Leu Ser Pro Gly Gly Arg Gly His Asn Glu Pro Lys Leu Cys  
 65 70 75 80

102

Gln Leu Asp Ser Ser Leu Gly Gln Gln Arg Lys Glu Leu Phe Thr Arg  
85 90 95

Lys Lys Lys Lys Thr Lys Lys Lys Lys Gly Gly Gly Gly Asn Thr  
100 105 110

Gly Ala Gln Thr Arg Gly Pro Gly Gly Asn Gly Gly Thr Arg Asp  
115 120 125

His Lys Phe Pro Lys Gln  
130

<210> 200  
<211> 34  
<212> PRT  
<213> Homo sapien

<400> 200

Met Tyr Pro Pro Gln Ala Leu Cys Glu Asn Ile His Glu Asp Tyr Ser  
1 5 10 15

Leu Ser Phe Tyr Thr Lys Arg Thr Thr Gln Arg Arg Pro Leu Gly Gly  
20 25 30

Phe Leu

<210> 201  
<211> 137  
<212> PRT  
<213> Homo sapien

<400> 201

Met Val Gly Arg Thr Thr Phe Tyr Lys Leu Arg Glu Ser Thr Gln Arg  
1 5 10 15

Ser Pro Leu Glu Arg Ala His Glu Glu Thr His Lys Ser Pro His Ala  
20 25 30

Val Cys Trp Leu Arg Glu Ile Asn Arg Ala Ser Ser Leu Leu Ser Leu  
35 40 45

Ser Leu Cys Val Gly Ala Arg Arg Ser Gln Thr Leu Cys Glu Lys Glu  
50 55 60

Lys Val Leu Ser Glu Arg Glu Ser Val Gly Val His Thr Glu Ser Gly  
65 70 75 80

Val Tyr Met Phe Tyr Ser Leu Trp Arg Val Ser Phe Ser Thr His Thr

103

85                      90                      95

Gly Ala His Asp Leu Ser His Lys Glu His Arg Thr His Thr Leu Trp  
                          100                      105                      110

Arg Ala Leu Ser His Leu Ile Phe Cys Glu Asn Val Lys Thr Phe Val  
                          115                      120                      125

Glu Arg Glu Val Phe Leu Pro Val Leu  
                          130                      135

<210> 202  
 <211> 134  
 <212> PRT  
 <213> Homo sapien

<400> 202

Met Val Val Arg Gln Tyr Val Ser Glu Ile Phe Glu Pro Ala Pro Pro  
   1                         5                         10                         15

Ser Thr Asn Lys His Tyr Phe Lys Arg Gly Lys Gly Ile Ser Met Glu  
                          20                         25                         30

Ala His Ser Arg Arg Gln Ser His Ser Leu Thr Arg Ser Ser Asp Pro  
                          35                         40                         45

Phe Ser Leu Gln His Arg Thr Gln Leu Leu Gln His Gly Ser His His  
                          50                         55                         60

His Gly Asp Leu Gly Pro Tyr Phe Ile Pro His Arg Met Glu Glu Ser  
   65                         70                         75                         80

Arg Leu Leu Leu Ser Leu Ser Ser Arg His Ser Phe Thr Ala Thr Phe  
                          85                         90                         95

Asp Gln Leu Leu Ala Arg Gly Lys Ala Ser Ser Thr Gly Thr Ser Arg  
                          100                         105                         110

Cys Pro Gly Leu Gly Ala Gly Ala Arg Arg Pro His Trp Ala Arg Val  
                          115                         120                         125

Ser Ser Ala Ala Thr Thr  
                          130

<210> 203  
 <211> 60  
 <212> PRT  
 <213> Homo sapien

104

&lt;400&gt; 203

Met Ile Ile Leu Cys Leu Ile Asn His Asn Ile Met Cys Trp Trp Val  
1 5 10 15

Ser Ser Ser Ser Asp Tyr Leu Ser Ile Ser Val Cys Val Val Gln Ile  
20 25 30

Ser Ser Arg Gly Val Ser Pro Cys Ala Arg Asp Lys Thr Thr Ala Leu  
35 40 45

Ser Leu Leu Ser Arg Ser Ser Leu Ser Tyr Leu Cys  
50 55 60

&lt;210&gt; 204

&lt;211&gt; 49

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 204

Met Asp Gly Thr Glu Gly Lys Gln Leu Phe Met Tyr Thr Ser Lys Arg  
1 5 10 15

Gly Lys Lys Lys Lys Lys Arg Asn Pro Leu Ile Ser Thr Leu Pro Ile  
20 25 30

Arg Gln Asp Ile Ser Thr Ser Gln Ile Leu Arg Phe Leu Ile Ser Arg  
35 40 45

Phe

&lt;210&gt; 205

&lt;211&gt; 53

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 205

Met Ser Pro Trp Leu Asn Glu Arg Ser Ile Ala Lys Tyr Leu Met Asp  
1 5 10 15

Lys Val Thr Thr Ala Leu Gln Ala Asn Asn His Ile Ser Pro Tyr Ile  
20 25 30

Asp Gln Gln Arg Tyr Tyr Asn Tyr Ala Ser Val Gly Ile Gln Pro Arg  
35 40 45

Leu Thr His Ile Thr  
50



105

<210> 206  
 <211> 219  
 <212> PRT  
 <213> Homo sapien

<400> 206

Met Thr Met Asn Thr Arg Ser Tyr Leu Thr Thr Phe Gly Ser Leu His  
 1 5 10 15

Ser Tyr Ser Ser Pro Gln Leu Trp Cys Asp Thr Leu Thr Leu Val Arg  
 20 25 30

His Gly Ser Ser Leu Gly His Asn Thr Arg Thr Asp Pro Thr Ala Tyr  
 35 40 45

Pro Ser Pro Tyr Cys Pro Tyr Leu Ala Glu His Phe Thr Leu Leu His  
 50 55 60

Lys Leu Ser Ser Met Thr Pro Gly Arg Leu Asp Met Ala Met Pro Tyr  
 65 70 75 80

Val Leu Ala Pro His Leu Ala Thr Pro Thr Pro Pro Ser Leu Thr Pro  
 85 90 95

Leu Arg Asn Asn Thr Thr Pro Ser His His His Thr Ile Thr Tyr Leu  
 100 105 110

Thr Thr Ala Pro Tyr His Arg Thr Leu Leu Thr Ser Pro Thr His Pro  
 115 120 125

Tyr Gly Asp Asp His Leu Tyr Leu Tyr Leu Thr Leu Thr Thr Pro Phe  
 130 135 140

Glu Pro Arg Pro Thr His Arg Tyr Pro Leu Pro Pro Leu Asn Pro Leu  
 145 150 155 160

Arg Ile Thr Thr Gln His Thr Ser Asp Gly Thr Thr Pro Phe Arg Asn  
 165 170 175

Thr His Pro Lys Leu His Pro Leu Tyr Tyr Thr Thr Gln His His Tyr  
 180 185 190

Tyr Tyr Ala His His Asn Gln Pro Gln Thr Ser Thr Thr Thr Ile Lys  
 195 200 205

His Ser Ala Gly Gln His Ser Glu Gln Gln Gln  
 210 215

106

<210> 207  
 <211> 97  
 <212> PRT  
 <213> Homo sapien

<400> 207

Met His Ala Arg Ala Ala Gln Cys Asp Gly Ser Ala Ala Gly Gln Val  
 1 5 10 15

Leu Pro Phe Phe Phe Phe Phe Phe Phe Phe Phe Leu Arg Gly Ser  
 20 25 30

Asn Leu Asp Pro Phe Phe Val Lys Lys Ile Phe Phe Phe Phe Phe  
 35 40 45

Phe Phe Leu Trp Lys Pro Pro Leu Glu Thr Ser Ala Ala Ala Leu Pro  
 50 55 60

Val Thr Thr Cys Leu Leu Ser Arg His Ser Cys Val Ile Gln Arg Asp  
 65 70 75 80

Gly Ala Pro Ala Gly Trp Lys Arg Glu Trp Pro Pro Arg Ala Gly Arg  
 85 90 95

Gly

<210> 208  
 <211> 261  
 <212> PRT  
 <213> Homo sapien

<400> 208

Met Leu Phe Cys Leu Pro Pro Arg Arg Ala Arg Val Cys Val Cys Cys  
 1 5 10 15

Ile Thr Leu Gly Gly His Ser Ser Leu Tyr Gly Lys Arg Cys Val Leu  
 20 25 30

Ser Leu Ala Arg Gly Arg Asp Ile Tyr Val Asn Thr Leu Ala Gly Glu  
 35 40 45

His Thr His Thr His Ser Tyr Ile Thr Gln Leu Phe Phe Val Cys Lys  
 50 55 60

Asn Met Phe Val Val His Leu Cys Val Cys Val Ile Trp Leu Tyr Thr  
 65 70 75 80

107

His Leu Ser Val Tyr Ile Leu Cys Val Cys Thr Arg Ala Ile Ala His  
85 90 95

Thr Leu Tyr Cys Pro Thr Ser Val Phe Met Arg Ala Arg Glu Arg Arg  
100 105 110

Gly Arg Val Arg Arg Glu Tyr Ile Ile Pro Thr Leu Cys Val Phe Ile  
115 120 125

Ile Thr Gln Leu Val Arg Glu Arg Glu His His Arg Arg Ser Ala Ala  
130 135 140

Val Cys Thr His Thr Arg His Thr Pro Leu Ser Leu Thr Pro Leu Leu  
145 150 155 160

Ser Tyr Ile His Thr Pro Arg Cys Ser Arg Arg Glu Tyr Ile Gly Cys  
165 170 175

Leu Tyr Ser Phe Thr His Phe Pro Val Gly Leu Tyr Ser His Thr Thr  
180 185 190

Ser Thr Ser Leu Leu Val Ser Thr His Thr His His Lys Ile Asn Thr  
195 200 205

Phe Leu Tyr Thr Pro Thr Leu Gln His Ser Leu Pro Pro His Leu Val  
210 215 220

Tyr Arg His Thr His Ser Leu Leu Pro Pro Pro Ala His Pro Gln Lys  
225 230 235 240

Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Gly Gly Asp  
245 250 255

Leu Arg Pro Ala Asp  
260

<210> 209  
<211> 111  
<212> PRT  
<213> Homo sapien

<400> 209

Met Arg Ser Thr His Trp Ala His Gly Thr Phe Leu Thr Pro Thr His  
1 5 10 15

Pro Phe Leu Ile Ser Ser Thr Phe Leu Ser Ile Tyr Leu Pro Pro Ala  
20 25 30

108

Pro Thr Pro Ile Pro Leu Ser Thr Thr Asn Pro Leu Ile Gln Ala Pro  
35 40 45

Pro Gly Pro Leu Ile Ile Lys Thr Ile Val Pro Leu Phe Leu Asn Met  
50 55 60

Asp Gln Lys Lys Lys Lys Asn Lys His Leu Ala Ala Thr Thr Ile  
65 70 75 80

His His Asn Ala Pro Leu Glu His Ala Ser Arg Tyr Thr Glu Ala Pro  
85 90 95

Ile Val Ile Ile His Ser Ser Phe Phe Leu Phe Phe Phe Val Phe  
100 105 110

<210> 210  
<211> 30  
<212> PRT  
<213> Homo sapien

<400> 210

Met Ala His Phe Ala Gln Gln Cys Ser Phe His Met Gln Leu Ile Thr  
1 5 10 15

His Asp Val Met Trp Ile Asp Thr Val Leu Thr Gln His Ile  
20 25 30